PATENT ABSTRACTS OF JAPAN

(11)Publication number : 07-200187
(43)Date of publication of application : 04.08.1995

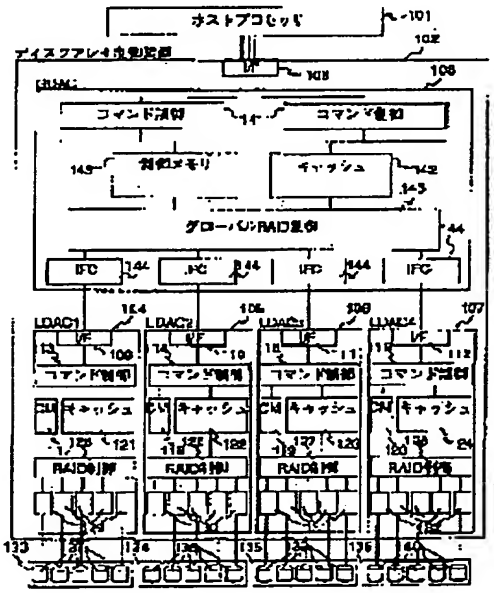
(51)Int.Cl. G06F 3/06
G06F 13/10

(21)Application number : 05-351130
(22)Date of filing : 30.12.1993

(71)Applicant : HITACHI LTD
(72)Inventor : TAKAMOTO YOSHIFUMI
TSUNODA HITOSHI

(54) DISK ARRAY DEVICE

(57)Abstract:
PURPOSE: To provide a disk array device which can carry out fast input/output operation, facilitates disk management, and causes no decrease in reliability even when many disk drives are connected.
CONSTITUTION: This disk array device is equipped with a global disk array controller 103 which puts local disk array controllers 104 to 107, where plural disk drivers 137 to 140 are connected, further in a disk array. Each disk array controller performs RAID control over plural disk drives in its machine. The global disk array controller performs RAID control over the local disk array controllers. Consequently, highly reliable disk devices are obtained by fastness by the division of data and global parity which are generated by the local disk arrays and global disk array and stored in the local disk arrays.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

(19)日本国特許庁（ＪＰ）　(12) **公開特許公報（A）**　(11)特許出願公開番号

特開平7－200187

(43)公開日　平成７年(1995)８月４日

(51)Int.Cl.[6]	識別記号	庁内整理番号	ＦＩ	技術表示箇所
Ｇ０６Ｆ　3/06	５４０			
13/10	３４０　Ａ	8327－5B		

審査請求　未請求　請求項の数13　ＦＤ　（全 28 頁）

(21)出願番号　特願平5－351130

(22)出願日　平成５年(1993)12月30日

(71)出願人　000005108
株式会社日立製作所
東京都千代田区神田駿河台四丁目６番地

(72)発明者　高本　良史
東京都国分寺市東恋ケ窪一丁目280番地
株式会社日立製作所中央研究所内

(72)発明者　角田　仁
東京都国分寺市東恋ケ窪一丁目280番地
株式会社日立製作所中央研究所内

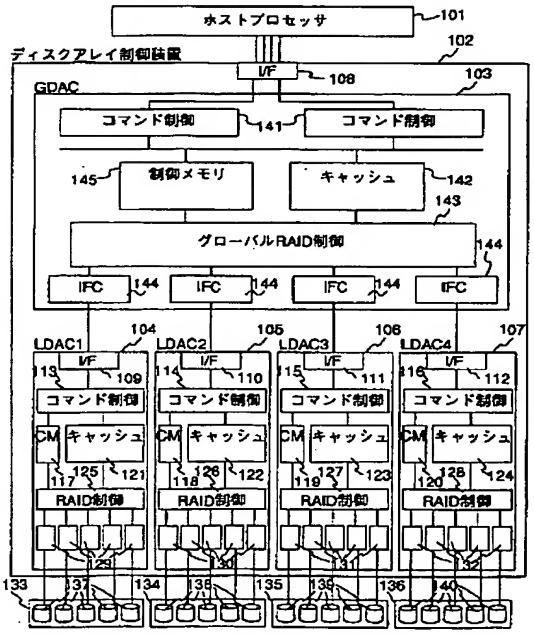
(74)代理人　弁理士　矢島　保夫

(54)【発明の名称】　ディスクアレイ装置

(57)【要約】　（修正有）

【目的】多数のディスクドライブを接続しても、高速な入出力が可能であり、またディスク管理も容易で信頼性を低下させることもないディスクアレイ装置を提供する。

【構成】複数のディスクドライブ１３７－１４０が接続された複数のローカルディスクアレイ制御装置１０４－１０７を、さらにディスクアレイ化するためのグローバルディスクアレイ制御装置１０３を備えている。ローカルディスクアレイ制御装置は、自機内の複数のディスクドライブをＲＡＩＤ制御する。グローバルディスクアレイ制御装置は、ローカルディスクアレイ制御装置に対しＲＡＩＤ制御する。この結果、データの分割による高速性と、ローカルディスクアレイとグローバルディスクアレイが生成しローカルディスクアレイに格納するグローバルパリティとにより高信頼なディスク装置が得られる。



1

【特許請求の範囲】

【請求項１】ホストプロセッサから転送されたデータを所定数のグループに分割し各グループのデータを並列に出力する手段と、

並列に出力された各グループのデータを、グループ単位でそれぞれ入力し格納する複数のローカルなディスクアレイ装置とを備えたことを特徴とするディスクアレイ装置。

【請求項２】複数のローカルなディスクアレイ装置と、

ホストプロセッサから転送されたデータを、前記複数のローカルなディスクアレイ装置のうちの何れに出力するかを選択し、選択したディスクアレイ装置に該データを出力する手段とを備えたことを特徴とするディスクアレイ装置。

【請求項３】ホストプロセッサから転送されたデータを所定数のグループに分割し各グループのデータを並列に出力する手段と、

前記複数のグループ間のパリティを生成して、前記各グループのデータと並列に出力する手段と、

並列に出力された各グループのデータを、グループ単位でそれぞれ入力し格納する複数のローカルなディスクアレイ装置と、

前記パリティを入力し格納するローカルなディスクアレイ装置とを備えたことを特徴とするディスクアレイ装置。

【請求項４】複数のディスクドライブを含み、ホストプロセッサから転送されたデータと前記データから生成されたパリティを、前記ディスクドライブに格納するディスクアレイ装置であって、

前記複数のディスクドライブが、それぞれ複数のディスクドライブを含む複数の論理グループに分けられているとともに、

前記ホストプロセッサから転送されたボリウム内アドレスとデータ長に基づいて、前記論理グループの一つまたは複数を選択し、選択した論理グループに、格納すべきデータを転送する手段と、

各論理グループごとに設けられ、前記転送されたデータから論理グループ内のパリティを生成し、自論理グループ内の複数のディスクドライブに前記データと生成したパリティとを格納する手段とを備えたことを特徴とするディスクアレイ装置。

【請求項５】複数のディスクドライブを含み、ホストプロセッサから転送されたデータと前記データから生成されたパリティを、前記ディスクドライブに格納するディスクアレイ装置であって、

前記複数のディスクドライブが、それぞれ複数のディスクドライブを含む複数の論理グループに分けられているとともに、

前記ホストプロセッサから転送されたボリウム内アドレスとデータ長に基づいて、前記論理グループの一つまた

2

は複数を選択し、選択した論理グループに、格納すべきデータを転送する手段と、

前記選択した各論理グループに転送するデータを用いて論理グループ間パリティを生成し、前記選択した論理グループとは異なる論理グループを選択し、選択した論理グループに、格納すべきデータとして前記論理グループ間パリティを転送する手段と、

各論理グループごとに設けられ、前記転送された格納すべきデータから論理グループ内のパリティを生成し、自論理グループ内の複数のディスクドライブに前記データと生成したパリティとを格納する手段とを備えたことを特徴とするディスクアレイ装置。

【請求項６】請求項４または５に記載のディスクアレイ装置であって、

前記論理グループ内のパリティが格納されたディスクドライブを除く他のディスクドライブに対し、各ディスクドライブへの格納単位ごとに順次連続した前記ボリウム内アドレスを割り付け、さらに前記ボリウム内アドレスは複数の前記論理グループに渡って順次水平に割り付けることを特徴とするディスクアレイ装置。

【請求項７】請求項６に記載のディスクアレイ装置であって、

前記論理グループにおける前記ボリウム内アドレスの連続部分を前記論理グループの最小格納単位とし、ホストプロセッサから転送されたデータを前記最小格納単位ごとに複数の前記論理グループに水平に分割し各論理グループに転送することを特徴とするディスクアレイ装置。

【請求項８】請求項７に記載のディスクアレイ装置であって、

前記各論理グループ内では、前記最小格納単位で転送されたデータを、ディスクドライブの格納単位ごとに水平に複数のディスクドライブに格納することを特徴とするディスクアレイ装置。

【請求項９】請求項８に記載のディスクアレイ装置であって、

前記論理グループの前記最小格納単位を満たしたデータから、順次、前記データに対応する前記論理グループに転送することを特徴とするディスクアレイ装置。

【請求項１０】請求項６に記載のディスクアレイ装置であって、

前記論理グループにおける前記ボリウム内アドレスの連続部分のデータが前記ホストプロセッサから転送されたとき、前記転送されたデータからパリティを生成し、前記データと前記パリティとを並列に前記論理グループ内のディスクドライブに格納することを特徴とするディスクアレイ装置。

【請求項１１】請求項４または５に記載のディスクアレイ装置であって、

前記論理グループ内の何れかのディスクドライブに障害が発生したとき、前記論理グループ内に格納されたパリ

ティと障害ディスクドライブ以外のディスクドライブのデータとから、前記障害ディスクドライブのデータを生成することを特徴とするディスクアレイ装置。

【請求項１２】請求項５に記載のディスクアレイ装置であって、

前記論理グループの何れかに障害が発生したとき、前記論理グループ間パリティと障害を起こした論理グループ以外の論理グループのデータとから、前記障害を起こした論理グループのデータを生成することを特徴とするディスクアレイ装置。

【請求項１３】複数のディスクドライブを含み、ホストプロセッサから転送されたデータと前記データから生成されたパリティを、前記ディスクドライブに格納するディスクアレイ装置であって、

前記複数のディスクドライブが、それぞれ複数のディスクドライブを含む複数の論理グループに分けられているとともに、各論理グループごとに論理グループ制御装置が設けられており、

該論理グループ制御装置は、ホストプロセッサから転送された入出力コマンドが、自装置に対応する論理グループに対する入出力要求かどうかを判定し、自装置に対応する論理グループに対する入出力要求である場合はその入出力コマンドを実行し、そうでなければその入出力コマンドをキャンセルする手段を備えたことを特徴とするディスクアレイ装置。

【発明の詳細な説明】

【０００１】

【産業上の利用分野】本発明は、コンピュータシステムなどに用いるディスクファイルシステムに関し、特に高速でかつ高信頼なディスクアレイ装置に関する。

【０００２】

【従来の技術】一般的に、コンピュータシステムは、プロセッサと２次記憶装置とを備えている。多く使用される２次記憶装置として、磁気ディスク装置がある。磁気ディスク装置において、現在その容量の伸び率は極めて高いが、メカニカル動作を伴う磁気ディスク装置の性能はプロセッサ性能の伸び率ほど高くない。その課題を解決する方式として、ディスクアレイが提案された。

【０００３】ディスクアレイに関する代表的な論文として、D.Patterson,G.Gibson,and R.H.Kartzらによる「A Case for Redundant Arrays of Inexpensive Disks(RAID),in ACM SIGMOD Conference,Chicago, IL, (June 1988)」がある。RAIDとは、複数のディスクドライブにデータを分散して配置することでアクセス時間を短縮し、かつ、誤り訂正用のパリティあるいはＥＣＣと呼ばれる冗長データを格納することで信頼性も高めることができる技術である。つまり、複数のディスクドライブに対して並列に入出力を行うことができることによる高速性と、何れかのディスクドライブに障害が発生したときでもパリティと障害ディスクドライブ以外のデータから障害ディスクドライブのデータを回復することができる技術である。

【０００４】この技術を実現するための特許出願もなされており、例えば、ジャイシヤンカー ムーセダス ミノン，ジェイムス マシューズ カーソン（インターナショナル ビジネス マシーンズ ＣＯＲＰ）らによる特開平４－２３０５１２号がある。この公報には、ディスクアレイにおける課題の一つであるパリティの更新を高速にする技術が開示されている。

【０００５】ディスクアレイでは、データ書き込み時に、データ自体とパリティの更新を行わなくてはならない。パリティの更新のためには、更新前のデータとパリティが必要となるケースがあり、そのためそれらのデータやパリティを事前に読みだすオーバヘッドと元のデータやパリティの位置に新しいデータやパリティを書き込むオーバヘッドがかかる。これをライト・ペナルティと呼ぶ。上記公報には、新しいデータやパリティを元の位置に記録せず、ディスクドライブの空き領域に格納することで、ライト・ペナルティによるオーバヘッドを少なくする技術が開示されている。

【０００６】他に、ディスクアレイの実装を容易にするための技術として、特開平５－４６５２４号に開示されたものがある。この公報に記載のディスクアレイ装置は、ディスクアレイの実装を容易にするために、複数のディスクを共通のバスに接続し、そのバスに接続された上位コントローラがディスクアレイ制御を行うものである。

【０００７】

【発明が解決しようとする課題】ディスクアレイコントローラの基本的な機能として、ホストプロセッサからの入出力要求を解釈／実行、データの分割、パリティの生成、キャッシュメモリ管理、およびディスクドライブの起動／終了処理がある。このように、ディスクアレイコントローラは、通常のディスクコントローラに比べて多くの処理を行なわなければならない。

【０００８】例えば、２台のドライブ（データを格納する１台のディスクドライブとパリティを格納する１台のディスクドライブ）でＲＡＩＤ５のディスクアレイ装置を構成したとする。このとき、新データの書き込み時には、旧データと旧パリティをそれぞれ２台のディスクドライブから読みだし、その後、新データと新パリティを前記２台のディスクドライブに出力しなければならない。つまり、１つの出力処理で４回のディスク入出力が発生することになり、これがライト・ペナルティになる。その他にも、ディスクアレイコントローラ特有の処理として、データの分割、パリティの生成などを行なう必要がある。

【０００９】従来のディスクアレイ装置は、一つのディスクアレイコントローラで複数のディスクドライブを制御していた。そのような構成で多数のディスクドライブ

5

を制御するためには、高速なディスクアレイコントローラを設けるか、同様のディスクアレイコントローラを複数設ける必要がある。

【００１０】しかし、高速なディスクアレイコントローラは非常に高価であり、これはディスクアレイ装置全体のコストを上げることにつながる。また、同様のディスクアレイコントローラを複数設けると、利用者から複数のディスク装置に見えてしまい、ディスク管理を複雑にする欠点がある。

【００１１】本発明の目的は、ディスクアレイ装置の改良にある。また、本発明の目的は、多数のディスクドライブを接続しても、高速な入出力が可能であり、コストを上げることがなく、またディスク管理も容易で信頼性を低下させることもないディスクアレイ装置を提供することにある。

【００１２】

【課題を解決するための手段】本発明では、複数のディスクドライブが、それぞれ複数のディスクドライブを含む複数の論理グループに分けられており、各論理グループは、例えば各論理グループごとに設けられたディスクアレイ制御装置によりそれぞれ管理される。そして、ホストプロセッサから転送されたボリウム内アドレスとデータ長に基づいて、前記論理グループの一つまたは複数を選択し、前記選択された論理グループごとに格納されるデータからパリティを生成し、前記論理グループ内のディスクドライブに前記データとパリティを格納するようにする。

【００１３】さらに、前記論理グループ内のパリティが格納されたドライブを除く他のディスクドライブに対し、ディスクドライブの格納単位ごとに順次連続した前記ボリウム内アドレスを割り付け、さらに前記ボリウム内アドレスは複数の前記論理グループに渡って順次水平に割り付ける。

【００１４】

【作用】複数のディスクドライブから構成される複数の論理グループに分け、前記論理グループごとに例えばディスクアレイ制御装置を設けて管理するので、小数のディスクドライブを管理する安価なディスクアレイコントローラになり、また、前記論理グループ内のパリティが格納されたドライブを除く他のディスクドライブに対し、ディスクドライブの格納単位毎に順次連続した前記ボリウム内アドレスを割り付け、さらに前記ボリウム内アドレスは複数の前記論理グループに渡って順次水平に割り付けることで、利用者からは単一のディスクドライブとして管理することができるようになる。

【００１５】

【実施例】以下、図面を用いて、本発明の実施例を詳細に説明する。

【００１６】［実施例１］図１は、本発明の第１の実施例に係るディスクアレイ装置の全体構成図を示したものである。

6

【００１７】ディスクアレイ制御装置（１０２）は、ホストプロセッサ（１０１）と接続されている。ディスクアレイ制御装置（１０２）は、複数のディスク装置（１３３，１３４，１３５，１３６）に接続されている。各ディスク装置（１３３，１３４，１３５，１３６）は、それぞれ複数のディスクドライブ（１３７，１３８，１３９，１４０）を備えている。

【００１８】ディスクアレイ制御装置（１０２）は、インタフェース制御部（１０８）と、グローバルディスクアレイ制御装置（１０３）と、複数のローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）とから構成されている。

【００１９】インタフェース制御部（１０８）は、ホスト（１０１）との間のデータあるいはコマンドの転送プロトコル制御を行う。グローバルディスクアレイ制御装置（１０３）は、複数のローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）に対して、データの分割、パリティの生成／格納、あるいはそれらの複数のローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）から転送されたデータのマージ処理などを行う。

【００２０】１つのローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）は、1つの論理グループに相当する。各ローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）は、グローバルディスクアレイ制御装置（１０３）から受けたデータの分割格納やパリティの生成／格納などを行う。論理グループとは、ＲＡＩＤ制御の単位である。すなわち、論理グループを構成するローカルディスクアレイ制御装置は、上位から受けたデータを分割しパリティを生成して複数のドライブに格納するが、それら複数のドライブを含むＲＡＩＤ制御の単位を論理グループという。

【００２１】グローバルディスクアレイ制御装置（１０３）と複数のローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）は共に、ディスク装置（１３３，１３４，１３５，１３６）の障害やディスク装置（１３３，１３４，１３５，１３６）内のディスクドライブ（１３７，１３８，１３９，１４０）の障害時に、パリティを使用して障害回復が可能である。

【００２２】本実施例の特徴の一つは、ローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）を制御するグローバルディスクアレイ制御装置（１０３）を設けていることである。これらの関連した詳細動作は後で説明する。

【００２３】グローバルディスクアレイ制御装置（１０３）は、コマンド制御部（１４１）、制御メモリ（１４５）、キャッシュメモリ（１４２）、グローバルＲＡＩＤ制御部（１４３）、およびインタフェースコントローラ（１４４）から構成されている。

7

【００２４】コマンド制御部（１４１）は、ホストプロセッサ（１０１）からの入出力要求を解釈し、制御メモリ（１４５）内のキューに入出力要求をキューイングする。制御メモリ（１４５）内には、ディスクアレイ装置全体を管理するための制御情報が格納されている。キャッシュメモリ（１４２）内には、ディスクアレイ装置（１０２）が入出力を行ったデータを、キャッシュメモリ（１４２）の容量の範囲内で格納しておく。頻繁にアクセスされるデータは、キャッシュメモリ（１４２）に残っている確率が高くなる。

【００２５】グローバルＲＡＩＤ制御部（１４３）は、複数のローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）を、ホスト１０１からは単体ディスクまたは複数ディスクとして取り扱うことができるような制御を行う。インタフェースコントローラ（１４４）は、ローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）との間のデータあるいはコマンドの転送制御を行う。

【００２６】以上が本実施例のディスクアレイ装置の構成概要であり、詳細な説明を以下に述べる。

【００２７】図２は、本発明における動作の概要の一例を示しており、本発明の特徴と効果の一例を以下に述べる。

【００２８】本動作概要は、ホストプロセッサ（１０１）からデータ（２２３）がディスク装置（１３３，１３４，１３５）内のディスクドライブに格納されるまでの動作とデータの流れを示している。本例は、ローカルディスクアレイ制御装置（１０４）がＲＡＩＤ３の動作を行なったケースである。

【００２９】ホストプロセッサ（１０１）から転送されたデータ（２２３）は、ディスクアレイ制御装置（１０２）内のグローバルディスクアレイ制御装置（１０３）で、ローカルディスクアレイ制御装置（１０４，１０５，１０６）の台数分（３つ）に分割される（２０１）。これらのデータ（２１９，２２０，２２１）は、並列にローカルディスクアレイ制御装置（１０４，１０５，１０６）に転送される。

【００３０】ローカルディスクアレイ制御装置（１０４，１０５，１０６）にデータが転送されると、ローカルディスクアレイ制御装置（１０４，１０５，１０６）内で、各ローカルディスクアレイ制御装置（１０４，１０５，１０６）に接続されているディスクドライブ数からパリティ格納用ディスクドライブ数を除いた台数分にデータの分割が行われ（２０７，２０９，２１１）、さらに同時にグローバルディスクアレイ制御装置（１０３）から転送されたデータのパリティが生成される（２０８，２１０，２１２）。ローカルディスクアレイ制御装置（１０４，１０５，１０６）内で分割されたデータとパリティは並列にディスクドライブに格納される。

【００３１】ローカルディスクアレイ制御装置（１０

8

４，１０５，１０６）は、それぞれ単独で動作可能な機構を持っているため、全てのディスクドライブの並列動作が可能となる。こういった構成により、大規模なデータの高速な入出力が可能となる。

【００３２】図３は、本発明における動作の概要の一例を示しており、本発明の特徴と効果の一例を以下に述べる。

【００３３】本動作概要は、ホストプロセッサ（１０１）からデータ（２２３）がディスク装置（１３３）内のディスクドライブに格納されるまでの動作とデータの流れを示している。本例は、ローカルディスクアレイ制御装置（１０４）がＲＡＩＤ５の動作を行なったケースである。

【００３４】ホストプロセッサ（１０１）からデータ（２２３）が転送されると、ディスクアレイ制御装置（１０２）内のグローバルディスクアレイ制御装置（１０３）では、そのデータを格納すべきローカルディスクアレイ制御装置（１０４）の選択が行なわれる。

【００３５】ローカルディスクアレイ制御装置（１０４）にデータが転送されると、ローカルディスクアレイ制御装置（１０４）内で、ローカルディスクアレイ制御装置（１０４）に接続されているディスクドライブ数からパリティ格納用ディスクドライブ数を除いた台数分にデータの分割が行われ（２０７）、さらに同時にグローバルディスクアレイ制御装置（１０３）から転送されたデータのパリティの一括生成が行なわれる（２０８）。

【００３６】通常ＲＡＩＤ５は、ブロック単位の入出力を行なうため、複数ブロックの格納時には各ブロック毎にライト・ペナルティが生じる。しかし、本例に示すように、連続して複数のブロックの書き込み時には、本発明によるように、連続してアドレスを設定することで、一括してパリティの生成を行なうことができるようになる。

【００３７】ローカルディスクアレイ制御装置（１０４）内で分割されたデータとパリティは並列にディスクドライブに格納される。こういった構成により、短いデータのスループットを向上することができる。

【００３８】図４は、本発明における動作の概要の一例を示しており、本発明の特徴と効果の一例を以下に述べる。

【００３９】本動作概要は、ホストプロセッサ（１０１）からデータ（２２３）がディスク装置（１３３，１３４，１３５，１３６）内のディスクドライブ（１３７，１３８，１３９，１４０）に格納されるまでの動作とデータの流れを示している。

【００４０】ホストプロセッサ（１０１）から転送されたデータ（２２３）は、ディスクアレイ制御装置（１０２）内のグローバルディスクアレイ制御装置（１０３）で、ローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）の台数からグローバルパリティ格

納用のローカルディスクアレイ制御装置（１０７）の台数を除いた数に分割される（２０１）。同時に、ホストプロセッサ（１０１）から転送されたデータ全体のグローバルパリティが生成される（２０２）。これらのデータは、並列にローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）に転送される。

【００４１】一般に、ディスクアレイは、信頼性の向上のために、データとは異なるパリティを生成しディスクドライブに格納しておく。もし、ディスクドライブに障害が発生しても、パリティと障害が発生したディスクドライブ以外のディスクドライブのデータとから、障害が発生したディスクドライブのデータを、演算により求めることができる。

【００４２】ローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）にデータが転送されると、各ローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）内で、ローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）に接続されているディスクドライブ（１３７，１３８，１３９，１４０）の数からパリティ格納用ディスクドライブを除いた台数分にデータの分割が行われる（２０７，２０９，２１１，２１３）。さらに同時に、グローバルディスクアレイ制御装置（１０３）から転送されたデータのパリティが生成される（２０８，２１０，２１２，２１４）。ローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）内で分割されたデータとパリティは、並列にディスクドライブ（１３７，１３８，１３９，１４０）に格納される。

【００４３】ローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）は、それぞれ単独で動作可能な機構を持っているため、全てのディスクドライブ（１３７，１３８，１３９，１４０）の並列動作が可能となる。こういった構成により、大規模なデータの高速な入出力が可能となる。

【００４４】さらには、通常のディスクアレイのパリティ以外に、ローカルディスクアレイ制御装置（１０４，１０５，１０６）間のパリティ（グローバルパリティ）を生成／格納していることから高い信頼性を得ることができる。つまり、各ローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）での回復処理と、上位のグローバルディスクアレイ制御装置（１０３）での回復処理といったように、本発明では障害回復処理が多層化されることによる。

【００４５】図５は、図２から図４で述べた特徴と効果とは別の特徴と効果の一例を示している。具体的には、本発明の入出力要求と終了報告の流れを示している。

【００４６】ホストプロセッサ（１０１）からデータが転送されると、グローバルディスクアレイ制御装置（１０３）はそのデータを分割しグローバルパリティを生成して、下位のローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）に並列に入出力を行なう（３０１）。ローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）は、グローバルディスクアレイ制御装置（１０３）から転送されたデータを分割しパリティを生成して、ディスクドライブ（１３７，１３８，１３９，１４０）に並列に入出力を行なう（３０６，３０７，３０８，３０９）。

【００４７】いま仮にローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）の数をｍとし、各ローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）に接続されたドライブ数をｄとすると、全部でｍｄ台のディスクドライブがあることになる。もし、一つのディスクアレイ制御装置にこのｍｄ台の全てのディスクドライブが接続されるとすると、ｍｄ回の入出力を一つの制御装置がまかなうことになり、非常に高速な制御装置が必要となる。

【００４８】しかし、本発明では、各ローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）毎にｄ回と少ない入出力回数を処理するだけでよく、ローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）に高速な処理は要求されない。

【００４９】図６は、コマンド制御部（１４１）の構成を示している。コマンド制御部（１４１）は、インタフェース制御部（１０８）とのコマンドやデータの転送制御を行うアダプタ（４０１）、プロセッサ（４０２）、ＤＭＡＣ（４０３）、メモリ（４０４）、バス（４０７）、およびバッファ（４０８）から構成されている。

【００５０】メモリ（４０４）内には、コマンド制御を行うマイクロプログラム（４０５，４０６）が格納されている。具体的には、コマンド受付処理プログラム（４０５）と終了処理プログラム（４０６）である。これらの詳細は後で述べる。

【００５１】マイクロプログラム（４０５，４０６）は、共通バス（４０８）を介して、プロセッサ（４０２）で実行される。コマンド制御部（１４１）と外部の制御部とのデータ転送は、ダイレクトメモリアクセスコントローラであるＤＭＡＣ（４０３）が、プロセッサ（４０２）とは独立に、実行する。コマンド制御部（１４１）から制御メモリ（１４５）やキャッシュにデータやコマンドを転送する際は、バッファ（４０８）とバス（４０９）を介して行われる。

【００５２】図７は、制御メモリ（１４５）に格納されているテーブルやリストを示している。５０２は仮想ドライブ管理テーブル、５０３はＩ／Ｏ管理リスト、５０４はローカルディスクアレイ管理テーブル、５０５はキャッシュ管理リスト、５０６はコマンドキューリストである。

【００５３】仮想ドライブ管理テーブル（５０２）は、複数のローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）の領域の単一制御やユーザの要求

に合わせて設定する領域を管理している。Ｉ／Ｏ管理リスト（５０３）は、ホストプロセッサ（１０１）から受け取った入出力要求を入出力が環境するまで入出力状態等と共に管理するテーブルである。

【００５４】ローカルディスクアレイ管理テーブル（５０４）は、グローバルディスクアレイ制御装置（１０３）に接続されているローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）を管理するテーブルであり、ＲＡＩＤレベル、容量、およびパーティション構成等が格納されている。キャッシュ管理リスト（５０５）は、キャッシュメモリ（１４２）内に格納されているデータとそのアドレスが格納されている。コマンドキューリスト（５０６）は、グローバルディスクアレイ制御装置（１０３）がコマンドに対してタグを付加する為の管理情報が格納されている。

【００５５】制御メモリ１４５のポインタ５０１からポイントされているこれらのテーブルやリストについての詳細な構成は後で述べる。

【００５６】図８は、ローカルディスクアレイ管理テーブル（５０４）の構成を示したものである。

【００５７】カラム６０１には、ローカルディスクアレイの識別子が格納されている。カラム６０２には、ローカルディスクアレイのパーティション識別子が格納されている。このカラムにより、単一のローカルディスクアレイ領域を複数に区切ることができる。これは、ディスクアレイが複数のディスクドライブから構成されていることから、単一領域として使用するには大きすぎることがある為である。このカラムはユーザが任意に指定可能である。

【００５８】カラム６０３にはパス識別子が格納されている。グローバルディスクアレイ制御装置（１０３）がローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）に対して入出力要求を発行するときに、このパス識別子を参照することで、ローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）の位置を知ることができる。カラム６０４は、各ローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）のドライブ保有数を示している。グローバルディスクアレイ制御装置（１０３）がグローバルパリティを作成する場合のパリティ位置を算出するときなどに使用する。

【００５９】カラム６０５は、各ローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）の制御しているＲＡＩＤレベルを示している。これもカラム６０４と同じように、グローバルパリティを作成する場合のパリティ位置を算出するときなどに使用する。カラム６０６には、各ローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）のパリティを格納するディスク数が格納されている。これもカラム６０５と同じように、グローバルパリティを作成する場合のパリティ位置を算出するときなどに使用する。

【００６０】カラム６０７には、各ローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）のパーティション毎のストライピングサイズが格納される。ストライピングサイズは、ローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）の最小入出力単位を意味する。例えば、ストライピングサイズが４（ＫＢ）では、４ＫＢ以内のデータ長のデータは、単一ディスクドライブに格納される。４ＫＢを超える長さのデータは、４ＫＢ毎に次のディスクドライブに分割して格納される。

【００６１】カラム６０８には、パーティション毎の容量が格納される。カラム６０９には、ステータスが格納されており、例えば障害の発生等の各ローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）の状態が格納される。

【００６２】図９は、仮想ドライブ管理テーブル（５０２）の構成を示している。仮想ドライブ管理テーブルは、ホストプロセッサ（１０１）から認識するディスクドライブと物理的なディスクドライブとのマッピングを行うために設けられる。

【００６３】カラム７０１には、仮想ドライブの識別子が格納される。カラム７０２から７０５までは、仮想ドライブ識別子で示される仮想ドライブが、ローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）のどのパーティションから構成されるかを示している。例えば、仮想ドライブＶＯＬ１は、ローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）のパーティション１から構成される。

【００６４】カラム７０６には、仮想ドライブ容量が格納される。カラム７０７には、グローバルＲＡＩＤレベルが格納されている。このグローバルＲＡＩＤレベルは、ローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）のＲＡＩＤレベルとは無関係に設定可能である。カラム７０８には、パリティを格納するディスク数が格納されている。カラム７０９には、グローバルディスクアレイ制御装置（１０３）のストライピングサイズが格納される。カラム７１０には、ステータスが格納されており、例えば障害の発生等の仮想ドライブ管理情報が格納される。

【００６５】図１０は、図７のような仮想ドライブ管理テーブル（５０２）を用いることにより、ホストプロセッサ（１０１）からどのように仮想ドライブが見えるか、および物理的にどのようにデータが配置されているかを示す概念図である。

【００６６】ホストプロセッサ（１０１）から入出力要求（８０１，８０２，８０３）が発行される。このとき、ホストプロセッサ（１０１）は、仮想ドライブ（８０４，８０５，８０６）に対して入出力を行なっている。このような仮想ドライブに対する入出力に対し、物

理的には、ディスク装置（１３３，１３４，１３５，１３６）にデータが分散配置される。ホストプロセッサ（１０１）から見える仮想ドライブ（８０４，８０５，８０６）は、物理的には、ＬＤＲＶ１（８０７），ＬＤＲＶ２（８０８），ＬＤＲＶ３（８０９）のように、ディスク装置（１３３，１３４，１３５，１３６）を横切る形態で分散されている。

【００６７】仮想ドライブ管理テーブル（５０２）の内容を変更することで、必ずしもディスク装置（１３３，１３４，１３５，１３６）の全てに渡って分散配置される構造にはならない。一つのローカルディスクアレイ領域を一つの仮想ドライブと定義することも可能である。さらに、８１０，８１１，８１２は、ローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）が生成／格納したパリティであり、８１３はグローバルディスクアレイ制御装置（１０３）が生成／格納したパリティを示している。こういったテーブルにより、ユーザの要求する容量を満足し、また、ＲＡＩＤレベル等を自由に設定することができるため、性能における柔軟性もある。

【００６８】図１１は、Ｉ／Ｏ管理リスト（５０３）の構成を示している。

【００６９】制御メモリ（１４５）からポイントされているアドレスは、さらに二つのポインタを示している。一つは次の入出力要求のリストが格納されている領域へのポインタ（９０１）であり、もう一つは入出力要求が格納されている領域を示すポインタ（９０２）である。９０９は、リストの最後であることを示す識別子である。

【００７０】格納されている入出力要求は、９０３から９０８の範囲で示されている。９０３は何処のホストプロセッサからの要求であるかを識別するためのホスト識別子、９０４は入出力を受け付けた時間、９０５は入出力の状態が格納されている。入出力の状態とは、例えば、入出力実行中などである。９０６は、ホストプロセッサから転送されたコマンドが格納されている。

【００７１】９０７は、ホストからコマンドが転送される際に付加されているタグが格納されている。このタグは、入出力を発行したホストプロセッサが、どの入出力要求が完了したのかを識別するためのものである。これは、ホストプロセッサが同一デバイスに対して複数の入出力要求を発行できる場合に付加される。もしこのタグがなければ、ホストプロセッサは、同一デバイスから終了通知が帰ってきても、どの入出力に対する完了かを識別することができないからである。

【００７２】９０８は新タグが格納されている。この新タグ（９０８）は、グローバルディスクアレイ制御装置（１０３）が設定するもので、ホストからのタグ（９０７）はホストが入出力を識別するものであるが、新タグ（９０８）はグローバルディスクアレイ制御装置（１０３）が同一ローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）に対して複数の入出力要求を発行したときに使用するものである。

【００７３】この新タグ（９０８）が必要な理由は、ホストが識別する仮想ドライブと、グローバルディスクアレイ制御装置（１０３）が識別するローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）とが同一でないことによる。つまり、タグを付加する側はドライブ毎にユニークなタグを付加する。しかし、このタグをそれを受け取った装置がホストの認識するドライブとは異なるドライブに使用すると、ドライブ毎のタグのユニークさが失われることになるためである。そのために、グローバルディスクアレイ制御装置（１０３）が使用するタグとしては、ホストから転送されたタグは使用せず、グローバルディスクアレイ制御装置（１０３）が管理できる新タグ（９０８）を使用する必要がある。

【００７４】図１２は、キャッシュ管理リスト（５０５）の構造を示している。キャッシュ管理リスト（５０５）は、キャッシュメモリ（１４２）内にキャッシングされているデータの場所やディスクドライブ（１３７，１３８，１３９，１４０）との対応等を管理するためのリストである。

【００７５】制御メモリ（１４５）からポイントされているアドレスは、さらに二つのポインタを示している。一つはキャッシュメモリのフリーポインタ（１００１）であり、もう一つは使用中ポインタ（１００２）である。キャッシュメモリ（１４２）は、ブロック単位で管理されており、前記フリーポインタ（１００１）と使用中ポインタ（１００２）が指すリストの一つ一つ（１００２－１００６，１００７－１０１１）は固定のブロックに対応している。

【００７６】フリーポインタ（１００１）からポイントされているリスト（１００２－１００６）は、キャッシュメモリ（１４２）の未使用領域（ブロック）を管理するために使用される。もし、新規にキャッシュメモリ（１４２）の領域を取得するときには、このフリーポインタ（１００１）からポイントされているリスト（１００２－１００６）の中から必要なだけのリストを取得した後、後述の使用中ポインタ（１００２）がポイントしているリスト内に入れる操作を行う。

【００７７】各リストには、次のリストをポイントするエリアと、該リストがディスク上のどこのデータ（ＢＬＫ＃）を保持しているかを示すエリアと、キャッシュメモリ（１４２）の何処に格納されているかを示すポインタとが、格納されている（１０４３－１０４５）。フリーポインタがポイントするリストは未使用リストがチェインされていることから、キャッシュメモリ（１４２）へのポインタの値は意味を持たない。

【００７８】１００７はハッシュリストである。ハッシュリスト（１００７）は、要求されたデータがキャッシ

15

ュメモリ（１４２）内に格納されているかどうかを高速に検索する目的で設けられている。要求するディスクドライブのブロック番号ＢＬＫ＃をキーとしてハッシングすれば、簡単に、検索しなければならない候補を絞ることが可能となる。

【００７９】図１３は、コマンドキューリスト（５０６）の構成を示している。

【００８０】１１０１は、各ローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）に対応したコマンドキューへのポインタが格納されている。１１０２と１１０３は、ローカルディスクアレイ制御装置（１０４）に対応するフリー新タグ（１１１０－１１１４）と使用新タグ（１１１５－１１１８）のリストをポイントしている。１１０４と１１０５は、ローカルディスクアレイ制御装置（１０５）に対応するフリー新タグ（１１１９－１１２２）と使用新タグ（１１２３－１１２７）のリストをポイントしている。１１０６と１１０７は、ローカルディスクアレイ制御装置（１０６）に対応するフリー新タグ（１１２８－１１３２）と使用新タグ（１１３３－１１３６）のリストをポイントしている。１１０８と１１０９は、ローカルディスクアレイ制御装置（１０７）に対応するフリー新タグ（１１３７－１１４１）と使用新タグ（１１４２－１１４５）のリストをポイントしている。

【００８１】各リストには、次のリストへのポインタと、新タグ値と、ホストプロセッサ（１０１）から転送された入出力コマンド（ホストからのタグも含む）とが、格納されている。フリー新タグ（１１１０－１１１４，１１１９－１１２２，１１２８－１１３２，１１３７－１１４１）中のコマンドエリア（ｃｍｄ）の値は意味を持たない。使用新タグのリストには、それぞれホストプロセッサ（１０１）からの入出力要求の一つが格納されている。使用新タグ（１１１５－１１１８，１１２３－１１２７，１１３３－１１３６，１１４２－１１４５）は、コマンドの実行終了時に、使用新タグリストからフリー新タグリストへ移される。各ローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）に対応するコマンドキューリストは、それぞれ重複しないタグ値を持っている。このコマンドキューリスト（５０６）により、前述のホストプロセッサ（１０１）から転送されたタグを新タグに変換することが可能となる。

【００８２】図１４は、コマンド受付処理プログラム（４０５）のフローチャートを示している。コマンド受付処理プログラム（４０５）は、コマンド制御部（１４５）内のメモリ（４０４）に格納されているマイクロプログラムである。

【００８３】ステップ１２０１ではキャッシュ確保可能かどうかをチェックする。もし可能であればステップ１２０２へ移り、そうでなければステップ１２１２へ移りコマンド異常終了処理を行う。通常、キャッシュメモリ

16

（１４２）に領域が確保できないことはない。コマンド異常終了処理が実行されるのは、キャッシュメモリ（１４５）がアクセス不可能な状態、例えば障害が発生したときである。

【００８４】ステップ１２０２では、キャッシュメモリ（１４５）の所要分領域を確保する。ステップ１２０３では、キャッシュ管理リスト（５０５）のフリーポインタ（１００１）からポイントされているリストから、所要分のリストを使用中ポインタ（１００２）がポイントしているリスト内に入れる。ステップ１２０４では、Ｉ／Ｏ管理リスト（５０３）中に空きリストが存在するかどうか調べる。もし、空きリストが存在すればステップ１２０５に進み、そうでなければステップ１２１２に進み異常終了処理を行う。この状態は、ディスクアレイ制御装置（１０２）内にホストプロセッサ（１０１）からのコマンドを受け付けるためのキューが無くなったことを意味する。

【００８５】ステップ１２０５では、Ｉ／Ｏ管理リスト（５０３）に新しいリストを追加する。これにより、当該コマンドがディスクアレイ制御装置（１０２）の処理対象となる。ステップ１２０６では、ホストＩＤをＩ／Ｏ管理リスト（５０３）のエリア（９０３）に格納する。ステップ１２０８では、コマンド受付時間をＩ／Ｏ管理リスト（５０３）のエリア（９０４）に格納する。この時間は、もしディスクアレイ制御装置（１０２）に障害が発生した場合などの保守に使用することができる。

【００８６】ステップ１２０９では、コマンドをＩ／Ｏ管理リスト（５０３）のエリア（９０６）に格納する。このとき、Ｉ／Ｏ管理リスト（５０３）のエリア９０７にもホストプロセッサ（１０１）から転送されたタグを格納する。これは、入出力要求完了時に、ホストプロセッサ（１０１）に対して完了情報と必要であればデータなどと共にホストプロセッサ（１０１）へ転送する際に参照される。ステップ１２１０では、ステップ１２０２で取得したキャッシュメモリ（１４２）の領域にコマンドを転送する。ここでコマンドとは、ＲＥＡＤ／ＷＲＩＴＥなどのディスク装置に対する指示と、ＷＲＩＴＥ要求であればデータも同様に、キャッシュメモリ（１４２）に格納される。

【００８７】ステップ１２１１では、Ｉ／Ｏ管理リスト（５０３）のステータスエリア（９０５）にコマンド受付完了を示すフラグを設定する。このステータスエリア（９０５）には、コマンド終了時には実行結果が格納されるが、それまでは、ディスクアレイ制御装置（１０２）内でのコマンドの実行状態がイベント毎に格納され、なんらかの障害発生時にロギング情報として使用することができる。

【００８８】図１５は、終了処理プログラム（４０６）のフローチャートを示している。終了処理プログラム

（４０６）は、コマンド制御部（１４５）内のメモリ（４０４）に格納されているマイクロプログラムである。

【００８９】ステップ１３０１では、要求されていたコマンドがＲＥＡＤ要求かＷＲＩＴＥ要求かを判定する。この判定は、Ｉ／Ｏ管理リスト（５０３）のコマンドエリアを参照することで達成可能である。判定の結果、ＲＥＡＤ要求であればステップ１３０６に進み、ＷＲＩＴＥ要求であればステップ１３０２へ進む。

【００９０】ステップ１３０２では、Ｉ／Ｏ管理リスト（５０３）のステータスエリア（９０５）とタグエリア（９０７）の内容をマージする。ステップ１３０３では、ステップ１３０２でマージしたデータを、Ｉ／Ｏ管理リスト（５０３）のホストエリア（９０３）が示すホストプロセッサ（１０１）へ、転送する。その後、ステップ１３０４に進む。ＷＲＩＴＥ処理はデータをすでにディスクドライブあるいはキャッシュエリアに書き込んだ後であるため、ホストプロセッサ（１０１）への情報は、処理が正常に終了したかどうかのみである。

【００９１】ステップ１３０６は、ＲＥＡＤ処理要求の終了時に実行される。キャッシュメモリ（１４２）内には、要求されたデータがすでに格納されている。ステップ１３０６では、このデータとステータスエリア（９０５）とタグエリア（９０７）の内容とをマージする。ステップ１３０７では、ステップ１３０６でマージしたデータを、Ｉ／Ｏ管理リスト（５０３）のホストエリア（９０３）が示すホストプロセッサ（１０１）へ、転送する。その後、ステップ１３０４に進む。

【００９２】ステップ１３０４では、Ｉ／Ｏ管理リスト（５０３）から、当該入出力要求を削除する。

【００９３】図１６は、グローバルＲＡＩＤ制御部（１４３）の構成を示している。グローバルＲＡＩＤ制御部（１４３）は、コマンドやデータの転送制御を行うアダプタ（１４０１）、プロセッサ（１４０２）、ＤＭＡＣ（１４０３）、メモリ（１４０４）、パススイッチ（１４０５）、およびパリティ生成部（１４０６）から構成されている。

【００９４】メモリ（１４０４）内には、コマンド制御を行うマイクロプログラム（１４０７，１４０８，１４０９，１４１０，１４１１，１４１２）が格納されている。具体的には、Ｉ／Ｏ要求受付処理プログラム（１４０７）、データ回復処理プログラム（１４０８）、Ｉ／Ｏ要求終了処理プログラム（１４０９）、データ配置制御プログラム（１４１０）、グローバルパリティ制御プログラム（１４１１）、およびタグ制御プログラム（１４１２）である。これらの詳細は後で述べる。マイクロプログラム（１４０７，１４０８，１４０９，１４１０，１４１１，１４１２）は、プロセッサ（１４０２）で実行される。

【００９５】外部の制御部とのデータ転送は、ダイレクトメモリアクセスコントローラであるＤＭＡＣ（１４０３）が、プロセッサ（１４０２）とは独立に実行する。パススイッチ（１４０５）は、ＤＭＡＣ（１４０３）から転送されたデータのヘッダを参照し、出力信号を振り分ける動作を行う。この動作については後で詳細に説明する。パリティ生成部（１４０６）は、パリティバッファ（１４１５）と、そのパリティバッファ（１４１５）に格納された（転送されてきた）データのパリティデータを作成するパリティジェネレータ（１４０８）とから構成される。

【００９６】グローバルＲＡＩＤ制御部（１４３）の大きな制御の流れは、以下のようなものである。まず、キャッシュメモリ（１４２）からデータを受け取り、データ配置制御プログラム（１４１０）によって必要であればデータの分割を行なう。次に、分割されたデータに、各々のデータをどのローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）に転送するかを識別するためのヘッダを付加する。その後、ヘッダに従いパススイッチ１４０５が各ローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）に分割後のデータを転送する。

【００９７】図１７は、Ｉ／Ｏ要求受付処理プログラム（１４０７）のフローチャートを示している。

【００９８】ステップ１５０１では、Ｉ／Ｏ管理リスト（５０３）のステータスエリア（９０５）に、グローバルＲＡＩＤ制御部（１４３）が処理要求を受け付けたことを示すフラグを、セットする。ステップ１５０２では、Ｉ／Ｏ管理リスト（５０３）のコマンドエリア（９０６）の内容を内部バッファに転送する。その後、ステップ１５０３に進み、データ配置制御プログラム（１４１０）に制御を移行する。

【００９９】図１８は、データ配置制御プログラム（１４１０）のフローチャートを示す。

【０１００】ステップ１６０１では、要求処理がＲＥＡＤかＷＲＩＴＥかを判定する。もしＲＥＡＤ処理要求であれば、ステップ１６０２に進み、要求データがキャッシュメモリ（１４２）に存在するかどうか判定する。そうでばければ、ステップ１６０３に進み、グローバルＲＡＩＤ制御部（１４３）がローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）に対して出力可能状態であるかどうか判定する。

【０１０１】ステップ１６０３について詳しく説明する。グローバルディスクアレイ制御装置（１０３）は、ローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）の上位に位置する制御装置であるため、ストライピングサイズが大きくなるために、ローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）よりもたくさんのデータをバッファリングした方が良いケースがある。例えば、４台のローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）が各

々４ＫＢのデータを４台のディスクドライブに１ＫＢづつ分割格納する構成で、グローバルディスクアレイ制御装置（１０３）が４台のローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）にフルストライピングするケースでは、グローバルディスクアレイ制御装置（１０３）は、

４ＫＢ＊４台＝１６ＫＢ

のデータをバッファリングすることによって４台のローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）に同時に出力処理ができる。この場合、ステップ１６０３では、全てのローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）に同時に出力可能かどうかを判定する。

【０１０２】しかし、このような判定が必要ないケースもある。上記の例では、フルストライピングを前提としたが、ＲＡＩＤ４，５のように、基本的にはデータをストライピングしない場合もある。その場合は、ホストプロセッサ（１０１）から転送されたデータを、バッファリングすることなく、いづれかのローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）に出力処理を行なう。また、別のケースとして、仮想ドライブが複数のローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）に渡って定義されていないときがある。この場合は、ローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）の通常の動作通りに、ローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）の分割動作に添って実行すれば良いので、グローバルディスクアレイ制御装置（１０３）はバッファリングする必要はない。従って、ステップ１６０３では、仮想ドライブ管理テーブル（５０２）やローカルディスクアレイ管理テーブル（５０４）を参照して、どのくらいのデータをバッファリングする必要があるかを判定する。

【０１０３】ステップ１６０５では、ホストコマンドのディスク要求アドレスを先頭ローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）のアドレスに変換する。この処理は、ホストプロセッサ（１０１）が認識しているドライブと、実際にローカルディスクアレイのドライブに格納される場所とが一致していないために行なう必要がある。参照するテーブルは、仮想ドライブ管理テーブル（５０２）とローカルディスクアレイ管理テーブル（５０４）である。

【０１０４】例えば、ホストプロセッサ（１０１）からの要求ディスクアドレスが、仮想ドライブ管理テーブル（５０２）のＶＯＬ２の先頭から３２ＫＢ目から１２ＫＢのデータ出力とする。その場合のローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）の先頭アドレスの求め方は、以下の（１）～（４）の通りである。

【０１０５】（１）図９の仮想ドライブ管理テーブル

（５０２）を参照して、ローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）にどのように仮想ドライブが配置されているかを求める。その結果、ＶＯＬ２はローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）のパーティション２から構成されていることがわかる。

【０１０６】（２）ローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）のパーティション２は、図８のローカルディスクアレイ管理テーブル（５０４）のカラム６０５，６０６，６０８から、それぞれ、ＲＡＩＤレベル３、パリティ数１であり、パーティション２は各ローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）の１ＧＢ目から始まることがわかる。

【０１０７】（３）図８のローカルディスクアレイ管理テーブル（５０４）のカラム６０４，６０７から、各ローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）のドライブ数は５台であり、ストライピングサイズは４ＫＢであることがわかる。パリティ数は１であることから、各ローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）には、４ＫＢ＊（５台－１台）＝１６ＫＢ格納されることになる。

【０１０８】（４）その結果、要求アドレスである３２ＫＢ目の物理的なアドレスは、３２ＫＢ／１６ＫＢ＋１＝３番目のローカルディスクアレイ制御装置（１０６）の１ＧＢ目が先頭アドレスとなる。

【０１０９】次に、ステップ１６０５では、要求コマンドがＲＥＡＤかＷＲＩＴＥかを判定する。その結果、ＲＥＡＤであればステップ１６０６に進み、ＷＲＩＴＥであればステップ１６１１に進む。

【０１１０】ステップ１６０６では、ローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）に対して入出力要求を発行するためにコマンドを生成する。このとき、各ローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）への入出力単位は、図９のカラム７０９のストライピングサイズを指定する。ステップ１６０７では、新タグを生成するためにタグ制御プログラム（１４１２）を実行する。タグ制御プログラム（１４１２）の詳細は後で述べる。

【０１１１】ステップ１６０８では、ステップ１６０６とステップ１６０７で生成したコマンドと新タグとをマージし、さらに、どのローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）への要求なのかを識別するためのヘッダを付加し、パススイッチ（１４０５）にデータ転送を行なう。ステップ１６０９では、次のローカルディスクアレイ制御装置を選択する。この操作は、上記物理アドレスを求める際に行なった計算と同じ要領で可能である。単に格納要求アドレスに図９のカラム７０９のストライピングサイズを加えたアドレスとして再計算するだけで可能となる。

【０１１２】次に、ステップ１６１０では、全てのデータを処理し終わったかどうかを判定し、まだ処理できていなければステップ１６０６から繰り返し、処理し終われば当処理を終了する。

【０１１３】ステップ１６０５でＷＲＩＴＥと判定されたときは、ステップ１６１１に進む。ステップ１６１１は、グローバルパリティ制御プログラムの実行を意味している。この処理についてはあとで詳細説明を行なう。

【０１１４】ステップ１６１２では、ローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）に対して入出力要求を発行するためにコマンドを生成する。このとき、各ローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）への入出力単位は、図９のカラム７０９のストライピングサイズを指定する。ステップ１６１３では、新タグを生成するためにタグ制御プログラム（１４１２）を実行する。タグ制御プログラム（１４１２）の詳細は後で述べる。

【０１１５】次に、ステップ１６１４では、ステップ１６０８と同様に、ローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）への入出力要求を発行する。ステップ１６１５はステップ１６０９と同等の処理であり、ステップ１６１６はステップ１６１０と同等の処理である。

【０１１６】図１９は、パススイッチ（１４０５）の動作を説明したものである。

【０１１７】この例では、４つのコマンド（１７０９，１７１０，１７１１，１７１２）がセレクタ（１４１４）によって分散転送される様子を示している。各コマンドは、ＲＥＡＤ／ＷＲＩＴＥなどのコマンドフィールド（１７０１，１７０３，１７０５，１７０７）と、転送先を示すパス番号（１７０２，１７０４，１７０６，１７０８）とから構成されている。

【０１１８】例えば、パケット１７１２はパス番号が１、パケット１７１１はパス番号が２、パケット１７１０はパス番号が３、パケット１７０９はパス番号が４である。セレクタ（１４１４）は、このパス番号を参照することによって、パス番号に続くデータをどのパスに転送すべきなのかを判定して、パスのスイッチングを行なう。例えば、パス番号が１であれば、セレクタ（１４１４）はパス番号１を選択し、データを転送する（１７１３）。同様に、パケット１７１１，１７１０，１７０９も、セレクタ（１４１４）によって、１７１４，１７１５，１７１６に分散して転送される。

【０１１９】図２０は、グローバルパリティ制御プログラム（１４１１）のフローチャートを示したものである。

【０１２０】ステップ１８０１では、仮想ドライブ管理テーブル（５０２）からパリティ位置の算出を行なう。例えば、要求ディスクアドレスが仮想ドライブ管理テーブル（５０２）のＶＯＬ２の先頭から３２ＫＢ目から１２ＫＢのデータ出力とする。その場合のパリティ位置の求め方は以下の（１）～（４）の通りである。

【０１２１】（１）仮想ドライブ管理テーブル（５０２）を参照して、ローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）にどのように仮想ドライブが配置されているかを求める。その結果、ＶＯＬ２はローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）のパーティション２から構成されていることがわかる。

【０１２２】（２）ローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）のパーティション２は、ローカルディスクアレイ管理テーブル（５０４）のカラム６０５，６０６，６０８から、それぞれ、ＲＡＩＤレベル３、パリティ数１であり、パーティション２は各ローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）の１ＧＢ目から始まることがわかる。

【０１２３】（３）グローバルＲＡＩＤレベルは３であることから、パリティは割当最終ローカルディスクアレイ（１０７）であることがわかる。

【０１２４】（４）その結果、パリティの物理的なアドレスは、ローカルディスクアレイ（１０７）の１ＧＢ目からとなる。

【０１２５】上記パリティ位置計算は一例であり、ローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）のＲＡＩＤレベルや、グローバルディスクアレイ制御装置（１０３）のＲＡＩＤレベルによって異なってくるが、各ＲＡＩＤレベルの配置規則を前述の演算に適用することで容易に実現可能である。

【０１２６】次に、ステップ１８０２では、パリティ演算に現在ドライブに格納されているデータが必要かどうかの判定を行なう。つまり、ＲＡＩＤレベルによっては、パリティ全体を更新できないことがある。例えばＲＡＩＤレベル５等である。パリティは、ＲＡＩＤレベル５に限らず複数のデータから一つあるいは一つ以上のパリティを生成する。このとき、あるパリティに関連する全てのデータが同時に更新されれば、その新しいすべてのデータから全く新しいパリティを生成できる。しかし、ＲＡＩＤレベル５等のように部分的なデータの更新である場合は、まず更新すべき箇所の過去のデータの情報をパリティから削除した後、新しいデータにおけるパリティを作成しなければならない。

【０１２７】ステップ１８０２の判定結果から旧データの必要がない場合はステップ１８０６に進み、そうでなければステップ１８０３に進む。ステップ１８０３では、旧データの読み込みのために、ローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）に対する入力コマンドを生成する。この処理は、図１８のステップ１６０６などと同等である。入力アドレスは、ステップ１８０１で求めたアドレスを使用する。ステップ

１８０４では、新タグを生成するためにタグ制御プログラム（１４１２）を実行する。この処理については、後で詳細に説明する。ステップ１８０５では、ローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）に対して入力要求を発行する。この処理は図１８のステップ１６０８等と同等である。

【０１２８】ステップ１８０６では、キャッシュ領域からパリティ生成部（１４０６）へ、グローバルディスクアレイ制御装置（１０３）のストライピングサイズを指定して、データ転送を行なう。必要であればステップ１８０５で取得した旧データも同時に転送する。グローバルディスクアレイ制御装置（１０３）のストライピングサイズは、図９のカラム７０９のストライピングサイズを参照することで求めることができる。この動作については後で詳細に説明する。

【０１２９】次に、ステップ１８０７では、パリティ生成部（１４０６）から、生成されたパリティを読み取る。ステップ１８０８では、ステップ１８０１で計算したパリティ位置に対して、ステップ１８０７で取得した新しいパリティを書き込むためのコマンドを生成する。ステップ１８０９では、ステップ１８０８で生成したコマンドに対するタグを取得するため、タグ制御プログラム（１４１２）を実行する。ステップ１８１０では、ステップ１８０１で求めたパリティ位置に新しいパリティを書き込むために出力処理を行なう。

【０１３０】図２１は、グローバルパリティの生成の様子を示している。

【０１３１】１９０１はパリティを生成するデータ群であり、１９０２はグローバルディスクアレイ制御装置（１０３）のストライピングサイズである。この２つのデータをパリティ生成部（１４０６）に転送することで、まずデータ群（１９０１）をグローバルディスクアレイ制御装置（１０３）のストライピングサイズに分割し（１９０３，１９０４，１９０５）、それぞれについて排他的論理和を取る（１９０６，１９０７）。これにより、グローバルパリティが生成される。

【０１３２】図２２は、ホストプロセッサ（１０１）から転送された入出力コマンドが、ディスク装置（１３３，１３４，１３５，１３６）に格納あるいは読みだしされるまでに、どのように変化するかを示している。

【０１３３】ホストプロセッサ（１０１）からは、データ（２００１）、データ長（２００２）、コマンド（２００３）、ホスト識別子（２００４）、タグ（２００５）、および仮想ドライブ識別子（２００６）が、一つのパケットとなって転送される。それを受けたグローバルディスクアレイ制御装置（１０３）は、ローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）に対して分割して入出力要求を発行するために、複数のコマンドを生成する。

【０１３４】その一つ一つは、２００７－２０１２で示

されるように、データ、データ長、コマンド、ホスト識別子、タグ、およびローカルディスクアレイ識別子（パス番号）が一つのパケットとなっている。ここで、パケットの形式は、２００１－２００６と同じであるが、内容は異なる。

【０１３５】例えば、データ（２００７）は分割されているためにデータ（２００１）に比べて少ない。それに伴い、データ長（２００８）もデータ長（２００２）より短い値がセットされている。コマンド（２００９）は、格納あるいは読みだし位置がローカルディスクアレイのアドレスに変わる。ホスト識別子（２０１０）は、ホストプロセッサ（１０１）の識別子とは異なり、この場合はグローバルディスクアレイ制御装置（１０３）の識別子である。タグ（２０１１）は、ホストプロセッサ（１０１）から転送されたタグではなく、新タグである。この理由は既に述べている。２０１２は、ホストプロセッサ（１０１）からは見えないローカルディスクアレイ識別子（パス番号）が格納されている。

【０１３６】ローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）内では、２０１３－２０１７に示すように、データ、データ長、コマンド、ホスト識別子、およびドライブ番号が一つのパケットとなってドライブに転送される。前述のグローバルディスクアレイ制御装置（１０３）内でのコマンドの変化と同じように、分割が行なわれる。

【０１３７】図２３は、タグ制御プログラム（１４１２）のフローチャートを示している。このタグは、前述の通り、ホストプロセッサ（１０１）から転送されたタグを、グローバルディスクアレイ制御装置（１０３）からローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）に対する入出力には使用できないことから、タグの変換が必要であることによる。

【０１３８】ステップ２１０１では、入出力を行なうローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）に対応する、コマンドキューリスト（５０６）中のフリータグポインタ（１１０２，１１０４，１１０６，１１０８）を検索する。フリータグポインタは、ローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）毎に設けられており、各ローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）に対する入出力に使用されていないタグがリスト形式でチェインされている（図１３）。ステップ２１０２では、このリストの中からタグを一つ取得する。

【０１３９】次に、ステップ２１０３では、取得したタグを使用中リストに繋ぎ換える。これにより、取得したタグは他の要求により使用されることが無くなり、同一ローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）内で重複するタグが使用されることはない。ステップ２１０４では、取得したタグを、Ｉ／Ｏ管理リスト（５０３）中の新タグエリア（９０８）に格納

25

する。これは、ホストプロセッサ（１０１）から転送されたタグと関連づけるために行なう。これにより、入出力完了時に、ホストプロセッサ（１０１）に対してホストプロセッサ（１０１）から転送されたタグを返送するとき、容易に検索することが可能となる。ステップ２１０５では、タグをコマンドに付加し、コマンドキューリスト（５０６）内のコマンド格納エリアに格納する。

【０１４０】図２４は、データ回復制御プログラム（１４０８）のフローチャートを示している。

【０１４１】ディスクアレイ装置にはデータの回復手段が設けられている。本発明では、通常のディスクアレイに比べより高信頼のディスクアレイにするために、パリティの階層化を行なっている。ローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）が管理するパリティと、グローバルディスクアレイ制御装置（１０３）が管理するグローバルパリティである。グローバルパリティは、複数のローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）に渡るパリティである。

【０１４２】ローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）が管理するパリティは、そのローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）内でのみ使用することができる。従って、ローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）内のいづれかのディスクドライブに障害が発生した場合に、そのパリティを使用して、障害ドライブのデータを回復することができる。

【０１４３】グローバルパリティは、複数のローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）に渡って作成されているため、前述のディスクドライブ単体の障害時にも使用することは可能であるが、それ以上にローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）そのものが障害を起こしたときに使用することができることの意味が大きい。

【０１４４】図２４のフローチャートでは障害回復を階層化した動作を示している。

【０１４５】ステップ２２０１では、ローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）の回復処理が可能かどうか判定する。この判定にはさまざまな要因が考えられる。例えば、障害回復機構の障害や、パリティで回復できる範囲を越えた障害時には、ローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）単体では障害回復ができない。この場合は、ステップ２２０２に進む。

【０１４６】ステップ２２０２では、パリティ領域の障害かどうか判定する。パリティ障害であれば、データ部分の障害ではないため、ここでは回復処理を行なわない。ステップ２２０３では、障害を起こしたローカルディスクアレイ以外のデータを取得する。この場合、パリティデータも含む。ステップ２２０４では、ステップ２

26

２０３で取得したデータあるいはパリティから、障害を起こした部分のデータを回復する。

【０１４７】なお、ここでは述べていないが、グローバルパリティのパリティも存在することがある。これは、グローバルディスクアレイ制御装置（１０３）がローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）にグローバルパリティデータを格納したときに、ローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）内では、グローバルパリティデータも通常のデータと同じくパリティを作成／格納するためである。本発明では、このようにパリティ情報も階層化されているため、高い信頼性を得ることができる。

【０１４８】図２５は、Ｉ／Ｏ要求終了処理プログラム（１４０９）のフローチャートを示している。

【０１４９】ステップ２３０１では、コマンドキューリスト（５０６）から終了する要求に対応するリストをサーチする。ステップ２３０２では、Ｉ／Ｏ管理リスト（５０３）からステップ２３０１と同様にサーチする。ステップ２３０４では、Ｉ／Ｏ管理リスト（５０３）のステータスエリア（９０５）に、ローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）から転送された終了状態を、格納する。

【０１５０】ステップ２３０４では、コマンドキューリスト（５０６）からステップ２３０１でサーチしたリストを削除し、未使用リストへ繋ぎ換える。これにより、本処理で使用したタグは、他の入出力のために使用可能となる。ステップ２３０５では、ホストプロセッサ（１０１）から転送されたタグをＩ／Ｏ管理リスト（５０３）のタグエリア（９０７）から取り出し、コマンド制御部（１４５）に通知することで処理を終了する。

【０１５１】次に、ローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）の詳細を述べる。この中で、コマンド制御部（１１３，１１４，１１５，１１６）は、グローバルディスクアレイ制御装置１０３のコマンド制御部１４１と同等である。また、インタフェース制御部（１０９，１１０，１１１，１１２）は、グローバルディスクアレイ制御装置１０３のインタフェース制御部１０８と同等である。また、インタフェースコントローラ（１２９，１３０，１３１，１３２）は、グローバルディスクアレイ制御装置１０３のインタフェースコントローラ１４４と同等である。

【０１５２】図２６は、ローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）のＲＡＩＤ制御部（１２５，１２６，１２７，１２８）の構成を示している。なお、図２６において、図１６に示したグローバルＲＡＩＤ制御部１４３の各部と共通の部分は、同じ番号で示している。

【０１５３】ＲＡＩＤ制御部（１２５，１２６，１２７，１２８）は、コマンドやデータの転送制御を行うアダプタ（１４０１）、プロセッサ（１４０２）、ＤＭＡ

27

C（１４０３）、メモリ（１４０４）、パススイッチ（１４０５）、およびパリティ生成部（１４０６）から構成されている。

【０１５４】メモリ（１４０４）内には、コマンド制御を行うマイクロプログラム（１４０７，２４０４，２４０１，２４０２，２４０３）が格納されている。具体的には、Ｉ／Ｏ要求受付処理プログラム（１４０７）、データ回復処理プログラム（２４０４）、Ｉ／Ｏ要求終了処理プログラム（２４０１）、データ配置制御プログラム（２４０３）、およびパリティ制御プログラム（２４０２）である。

【０１５５】Ｉ／Ｏ要求受付処理プログラム（１４０７）は、図１７と同様であるので説明を省略する。Ｉ／Ｏ要求終了処理プログラム（２４０１）、データ配置制御プログラム（２４０３）、パリティ制御プログラム（２４０２）、およびデータ回復処理プログラム（２４０４）の詳細は後で述べる。マイクロプログラム（１４０７，２４０４，２４０１，２４０２，２４０３）はプロセッサ（１４０２）で実行される。

【０１５６】外部の制御部とのデータ転送は、ＤＭＡＣ（１４０３）がプロセッサ（１４０２）とは独立に実行する。パススイッチ（１４０５）は、ＤＭＡＣ（１４０３）から転送されたデータのヘッダを参照し、出力信号を振り分ける動作を行う。パリティ生成部（１６０６）は、パリティバッファ（１４１５）と、そのパリティバッファ（１４１５）に格納された（転送されてきた）データのパリティデータを作成するパリティジェネレータ（１４０８）とから構成される。

【０１５７】ＲＡＩＤ制御部（１２５，１２６，１２７，１２８）の大きな制御の流れは、以下のようなものである。まず、グローバルディスクアレイ制御装置（１０３）からデータを受け取り、データ配置制御プログラム（２４０３）によって必要であればデータの分割を行なう。次に、分割されたデータに、各々のデータをどのディスクドライブ（１３７，１３８，１３９，１４０）に転送するかを識別するためのヘッダを付加する。その後、ヘッダに従いパススイッチ（１４０５）が各ディスクドライブ（１３７，１３８，１３９，１４０）に分割後のデータを転送する。

【０１５８】図２７は、制御メモリ（１１７，１１８，１１９，１２０）に格納されているリストを示している。２５０２はＩ／Ｏ管理リスト、２５０３はキャッシュ管理リストである。

【０１５９】Ｉ／Ｏ管理リスト（２５０２）は、グローバルディスクアレイ制御装置（１０３）から受け取った入出力要求を入出力が環境するまで入出力状態等と共に管理するテーブルである。キャッシュ管理リスト（２５０３）は、キャッシュメモリ（１２１，１２２，１２３，１２４）内に格納されているデータとそのアドレスが格納されている。Ｉ／Ｏ管理リスト（２５０２）とキャッシュ管理リスト（２５０３）は、図７のＩ／Ｏ管理リスト５０３（図１１）とキャッシュ管理リスト５０５（図１２）と同等のリストである。

28

【０１６０】図２８は、Ｉ／Ｏ要求終了処理プログラム（２４０１）のフローチャートを示している。

【０１６１】ステップ２６０１では、Ｉ／Ｏ管理リスト（２５０２）から終了する入出力要求のリストをサーチする。ステップ２６０２では、Ｉ／Ｏ管理リスト（２５０２）のステータスエリアにディスク装置（１３３，１３４，１３５，１３６）から転送された終了状態を格納することで処理を終了する。

【０１６２】図２９は、データ配置制御プログラム（２４０３）のフローチャートを示す。

【０１６３】ステップ２７０１では、要求処理がＲＥＡＤかＷＲＩＴＥかを判定する。もし、ＲＥＡＤ処理要求であれば、ステップ２７０２に進み、要求データがキャッシュメモリ（１４２）に存在するかどうか判定する。ステップ２７０１でＷＲＩＴＥ処理要求のときは、ステップ２７０３に進み、ＲＡＩＤ制御部（１２５，１２６，１２７，１２８）がディスク装置（１３３，１３４，１３５，１３６）に対して出力可能状態であるかどうか判定する。

【０１６４】ステップ２７０４では、グローバルディスクアレイ制御装置（１０３）のディスク要求アドレスを先頭ディスク装置（１３３，１３４，１３５，１３６）のアドレスに変換する。ステップ２７０５では、要求コマンドがＲＥＡＤかＷＲＩＴＥかを判定する。その結果、ＲＥＡＤであればステップ２７０６に進み、ＷＲＩＴＥであればステップ２７０９に進む。

【０１６５】ステップ２７０６では、ディスクドライブ（１３７，１３８，１３９，１４０）に対して入出力要求を発行する。ステップ２７０７では、次のディスクドライブ（１３７，１３８，１３９，１４０）を選択する。ステップ２７０８では全てのデータ処理が終了したかどうか判定し、終了していなければステップ２７０６から繰り返す。

【０１６６】ステップ２７０９はステップ２７０６と同等である。ステップ２７１０はステップ２７０７と同等である。ステップ２７１１はステップ２７０８と同等である。ステップ２７１２では、パリティ制御プログラム（２４０２）を実行する。

【０１６７】図３０は、パリティ制御プログラム（２４０２）のフローチャートを示したものである。

【０１６８】ステップ２８０１では、ＲＡＩＤレベル、ドライブ数、および要求ディスクアドレスからパリティ位置の算出を行なう。ステップ２８０２では、パリティ更新のために旧データが必要かどうか判定する。ステップ２８０２の判定結果から、旧データの必要がない場合はステップ２８０４に進み、そうでなければステップ２８０３に進む。

29

【０１６９】ステップ２８０３では、旧データの読み込みのために、ディスクドライブ（１３７，１３８，１３９，１４０）に対して旧データの入力要求を発行する。入力アドレスは、ステップ２８０１で求めたアドレスを使用する。

【０１７０】ステップ２８０４では、キャッシュ領域からパリティ生成部（１４０６）へデータ転送を行なう。必要であればステップ２８０３で取得した旧データも同時に転送する。ステップ２８０５では、パリティ生成部（１４０６）から、生成されたパリティを読み取る。ステップ２８０６では、ステップ２８０１で求めたパリティ位置に新しいパリティを書き込むために出力処理を行なう。

【０１７１】図３１は、データ回復制御プログラム（２４０４）のフローチャートを示している。ディスクアレイ装置にはデータの回復手段が設けられている。従って、ローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）内のいづれかのディスクドライブに障害が発生した場合に、そのパリティを使用して、障害ドライブのデータを回復することができる。

【０１７２】まず、ステップ２９０１では、パリティ領域の障害かどうか判定する。パリティ障害であれば、データ部分の障害ではないため、ここでは回復処理を行なわない。ステップ２９０２では、障害を起こしたディスクドライブ以外のデータを取得する。この場合、パリティデータも含む。ステップ２９０３では、ステップ２９０２で取得したデータあるいはパリティから、障害を起こした部分のデータを回復する。

【０１７３】図３２は、コマンド受付処理プログラムのフローチャートを示している。コマンド受付処理プログラムは、コマンド制御部（１１３，１１４，１１５，１１６）内のメモリに格納されているマイクロプログラムである。

【０１７４】ステップ３００１では、キャッシュ確保可能かどうかをチェックする。もし可能であればステップ３００２へ移り、そうでなければステップ３０１３へ移りコマンド異常終了処理を行う。ステップ３００２では、キャッシュメモリ（１２１，１２２，１２３，１２４）の所要分領域を確保する。ステップ３００３では、キャッシュ管理リスト（２５０３）のフリーポインタ（１００１）でポイントされているリストから、使用中ポインタ（１００２）がポイントしているリスト内に入れる。

【０１７５】ステップ３００４では、Ｉ／Ｏ管理リスト（２５０２）中に空きリストが存在するかどうか調べる。もし、空きリストが存在すればステップ３００５に進み、そうでなければステップ３０１３に進み異常終了処理を行う。

【０１７６】ステップ３００５では、新しいＩ／Ｏ管理リスト（２５０２）を追加する。ステップ３００６では、ホストＩＤをＩ／Ｏ管理リスト（２５０２）のエリア（９０３）に格納する。ステップ３００８では、コマンド受付時間をＩ／Ｏ管理リスト（２５０２）のエリア（９０４）に格納する。この時間は、もしディスクアレイ制御装置（１０２）に障害が発生した場合などの保守に使用することができる。

30

【０１７７】ステップ３００９では、コマンドをＩ／Ｏ管理リスト（２５０２）のエリア（９０６）に格納する。ステップ３０１０では、ステップ３００２で取得したキャッシュメモリ（１２１，１２２，１２３，１２４）領域にコマンドを転送する。ここでコマンドとはＲＥＡＤ／ＷＲＩＴＥなどのディスク装置に対する指示と、ＷＲＩＴＥ要求であればデータも同様にキャッシュメモリ（１４２）に格納される。

【０１７８】ステップ３０１１では、Ｉ／Ｏ管理リスト（２５０２）のステータスエリア（９０５）にコマンド受付完了を示すフラグを設定する。このステータスエリア（９０５）は、コマンド終了時には実行結果が格納されるが、それまでは、ローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）内でのコマンドの実行状態がイベント毎に格納され、なんらかの障害発生時にロギング情報として使用することができる。

【０１７９】図３３は、終了処理プログラムのフローチャートを示している。終了処理プログラムは、コマンド制御部（１１３，１１４，１１５，１１６）内のメモリに格納されているマイクロプログラムである。

【０１８０】ステップ１３０１では、要求されていたコマンドがＲＥＡＤ要求かＷＲＩＴＥ要求かを判定する。この判定は、Ｉ／Ｏ管理リスト（２５０２）のコマンドエリアを参照することで達成可能である。判定の結果、ＲＥＡＤ要求であればステップ３１０３に進み、ＷＲＩＴＥ要求であればステップ３１０２へ進む。

【０１８１】ステップ３１０２では、Ｉ／Ｏ管理リスト（２５０２）のステータスエリア（９０５）をグローバルディスクアレイ制御装置（１０３）へ転送する。その後、ステップ３１０５に進む。

【０１８２】ステップ３１０３は、ＲＥＡＤ処理要求の終了時に実行される。キャッシュメモリ（１２１，１２２，１２３，１２４）内には、要求されたデータがすでに格納されている。ステップ３１０３では、このデータとステータスエリア（９０５）の内容をマージする。ステップ３１０４では、ステップ３１０３でマージしたデータをＩ／Ｏ管理リスト（２５０２）のホストエリア（９０３）が示すグローバルディスクアレイ制御装置（１０３）へ転送する。その後、ステップ３１０５に進む。

【０１８３】ステップ３１０５では、Ｉ／Ｏ管理リスト（２５０２）から、当該入出力要求を削除する。

【０１８４】図３４は、本実施例における障害回復動作の一例を示す。この図を参照して、本発明による障害回

復動作の一例を説明する。

【０１８５】本例では、ローカルディスクアレイ制御部（１０４）が管理するディスクドライブの一つに障害が発生し、さらに、ローカルディスクアレイ制御装置（１０５）に障害が発生したとき、ホストプロセッサ（１０１）からデータの読みだし要求が出されたときの回復動作を示している。

【０１８６】まず、ローカルディスクアレイ制御装置（１０４）は、障害が発生したドライブ以外のドライブからデータを読みだし、障害が発生したドライブのデータを修復する（３２０４）。修復されたデータと正常ドライブのデータはマージされて（３２０３）、グローバルディスクアレイ制御装置（１０３）へ転送する。

【０１８７】ローカルディスクアレイ制御装置（１０５）は、要求されたデータ（３２０５）をそのままグローバルディスクアレイ制御装置（１０３）へ転送する。グローバルディスクアレイ制御装置（１０３）は、ローカルディスクアレイ制御装置（１０５）が単独で修復不可能であることを判断し、グローバルパリティを使用したデータ修復を行なう。

【０１８８】そのために、グローバルパリティが格納されているローカルディスクアレイ制御装置（１０７）から、回復に必要なグローバルパリティ（３２０６）を読みだす。グローバルパリティ（３２０６）と正常にデータが読みだされたローカルディスクアレイのデータは、グローバルディスクアレイ制御装置（１０３）内で、障害が発生したローカルディスクアレイ制御装置（１０５）のデータを修復する（３２０２）。修復されたデータは、正常にデータが読みだされたローカルディスクアレイのデータとマージされ（３２０１）、ホストプロセッサ（１０１）へ転送される。

【０１８９】このように、本発明では、ドライブ障害だけでなく、ディスクアレイそのものに障害が発生しても、ホストプロセッサ（１０１）の処理を中断することがない。

【０１９０】図３５は、本発明によるディスクアレイ装置のパリティデータの格納の一例を示している。図中、Ｐはローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）により生成されたパリティを示し、ＧＰはグローバルディスクアレイ制御装置（１０３）により生成されたパリティを示す。

【０１９１】３３０１－３３０４は、ローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）のＲＡＩＤレベルが３または４のときのパリティを示す。それに対し、３３００はグローバルディスクアレイ制御装置（１０３）が生成したパリティであり、このときグローバルディスクアレイ制御装置（１０３）のＲＡＩＤレベルは３または４である。

【０１９２】同様に、３３０９－３３１２は、ローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）のＲＡＩＤレベルが３または４のときのパリティを示す。３３０５－３３０８はグローバルディスクアレイ制御装置（１０３）が生成したパリティであり、このときグローバルディスクアレイ制御装置（１０３）のＲＡＩＤレベルは５である。このように、データのみならずパリティも複数のローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）に渡って格納される。

【０１９３】３３１３のグローバルパリティも同様である。このとき、ローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）はＲＡＩＤレベルが５である。

【０１９４】図３６は、図３５と同様にパリティデータの格納の一例を示している。図３５では、グローバルディスクアレイ制御装置（１０３）とローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）の両方でパリティを生成／格納することで、非常に高い信頼性を得られる例を示した。図３６は、信頼性よりも大容量を優先する使用方法を示した例である。

【０１９５】３４０１は、当ディスクアレイ装置の中で唯一つのパリティである。他のローカルディスクアレイ制御装置（１０４，１０５，１０６）にはパリティは存在しない。こうすることで、容量を優先するディスクアレイを構成することができる。

【０１９６】これと同様の別の例として、ローカルディスクアレイ制御装置（１０７）のすべてのディスクドライブにグローバルパリティを格納することもできる。３４０２は、ＲＡＩＤ５のグローバルパリティである。しかし、他のローカルディスクアレイ制御装置（１０４，１０５，１０６）にはパリティは存在しない。

【０１９７】同じＲＡＩＤ５のグローバルパリティでも、３４０３－３４０６のようにローカルディスクアレイ制御装置（１０４，１０５，１０６，１０７）に分散して配置することも可能である。

【０１９８】［実施例２］次に、本発明の第２の実施例を説明する。

【０１９９】図３７は、本発明の第２の実施例の全体構成図を示したものである。図１と共通の部分は同じ番号で示した。図１と大きく異なるところは、グローバルディスクアレイ制御装置（３５１３）とローカルディスクアレイ制御装置（３５０２，３５０３，３５０４，３４０５）との接続に、共通バス（３５０６）を用いたところである。制御メモリ（３５１１）、グローバルＲＡＩＤ制御部（３５０１）、およびコマンド制御部（３５０７－３５１０）については後で詳細説明を行なうが、その他の部分については図１と同等であるので説明を省略する。

【０２００】図３８は、グローバルＲＡＩＤ制御部（３５０１）の構成を示している。基本的な構成は、図１６と同じであるが、マイクロプログラム（３６０１）とマ

ルチキャスト制御部（３６０２）が異なる。

【０２０１】グローバルＲＡＩＤ制御部（３５０１）は、ホストプロセッサ（１０１）から入出力要求を受け取り、データをローカルディスクアレイ制御装置（３５０２，３５０３，３５０４，３４０５）に対して転送する。このとき、グローバルＲＡＩＤ制御部（３５０１）は、マルチキャスト制御部（３６０２）により共通バス（３５０６）に接続された全てのローカルディスクアレイ制御装置（３５０２，３５０３，３５０４，３４０５）を転送先として同一のコマンドを転送する。

【０２０２】図３９は、ローカルディスクアレイ制御装置（３５０２，３５０３，３５０４，３４０５）内の制御メモリ（３５１１）の構造を示している。

【０２０３】テーブルやリスト（５０２，５０３，５０４，５０５）は、図７と同等であるが、ローカルディスクアレイ識別子（３７０２）が追加されている。ローカルディスクアレイ識別子（３７０２）は、ローカルディスクアレイ制御装置（３５０２，３５０３，３５０４，３４０５）毎に異なる識別子が設定される。ローカルディスクアレイ識別子（３７０２）は、自ローカルディスクアレイ制御装置（３５０２，３５０３，３５０４，３４０５）がグローバルＲＡＩＤ制御部（３５０１）から転送されたコマンドやデータを実行すべきかどうかを判断するために使用する。

【０２０４】図４０は、コマンド受付処理プログラムのフローチャートを示している。コマンド受付処理プログラムは、コマンド制御部（３５０７，３５０８，３５０９，３５１０）内のメモリに格納されているマイクロプログラムである。

【０２０５】ステップ３８０１では、制御メモリ（３５１１）内のローカルディスクアレイ識別子（３７０２）と転送されたコマンドとから、自ローカルディスクアレイが実行すべき内容かどうかを判定する。この判定は、ローカルディスクアレイ識別子（３７０２）と仮想ドライブ管理テーブル（５０２）を用いて行なわれる。例えば、グローバルＲＡＩＤ制御部（３５０１）から仮想ドライブｎのブロックｍに対する入出力要求があった場合は以下のように判定する。

【０２０６】（１）仮想ドライブ管理テーブル（５０２）から仮想ドライブｎ、ブロックｍが格納されているローカルディスクアレイの識別子を得る。

（２）得られた識別子と制御メモリ（３５１１）内のローカルディスクアレイ識別子（３７０２）とを比較し、自ローカルディスクアレイが実行すべきかどうか判定する。

【０２０７】以上の動作により、ステップ３８０１の判定が可能となる。自ローカルディスクアレイで実行すべきコマンドであったときはステップ３８０２に進み、総出内ときは処理を終了する。

【０２０８】ステップ３８０２では、キャッシュ確保可

能かどうかをチェックする。もし可能であればステップ３８０３へ移り、そうでなければステップ３８１４へ移りコマンド異常終了処理を行う。ステップ３８０３では、キャッシュメモリ（１２１－１２４）の所要分領域を確保する。ステップ３８０４では、キャッシュ管理リスト（５０５）のフリーポインタ（１００１）でポイントされているリストから、所要分のリストを、使用中ポインタ（１００２）がポイントしているリスト内に入れる。

【０２０９】次に、ステップ３８０５では、Ｉ／Ｏ管理リスト（５０３）中に空きリストが存在するかどうか調べる。もし、空きリストが存在すればステップ３８０６に進み、そうでなければステップ３８１４に進み異常終了処理を行う。

【０２１０】ステップ３８０６では、Ｉ／Ｏ管理リスト（５０３）に新しいリストを追加する。ステップ３８０７では、ホストＩＤをＩ／Ｏ管理リスト（５０３）のエリア（９０３）に格納する。ステップ３８０９では、コマンド受付時間をＩ／Ｏ管理リスト（５０３）のエリア（９０４）に格納する。この時間は、もしディスクアレイ制御装置（１０２）に障害が発生した場合などの保守に使用することができる。ステップ３８１０では、コマンドをＩ／Ｏ管理リスト（５０３）のエリア（９０６）に格納する。

【０２１１】ステップ３８１１では、ステップ３８０３で取得したキャッシュメモリ（１２１－１２４）領域にコマンドを転送する。ここでコマンドとはＲＥＡＤ／ＷＲＩＴＥなどのディスク装置に対する指示と、ＷＲＩＴＥ要求であればデータも同様にキャッシュメモリ（１２１－１２４）に格納される。ステップ３８１２では、Ｉ／Ｏ管理リスト（５０３）のステータスエリア（９０５）にコマンド受付完了を示すフラグを設定する。このステータスエリア（９０５）には、コマンド終了時には実行結果が格納されるが、それまでは、ディスクアレイ制御装置（１０２）内でのコマンドの実行状態がイベント毎に格納され、なんらかの障害発生時にロギング情報として使用することができる。

【０２１２】図４１は、データ配置制御プログラム（１４１０）の動作フローチャートを示している。

【０２１３】ステップ３９０１では、要求処理がＲＥＡＤかＷＲＩＴＥかを判定する。もし、ＲＥＡＤ処理要求であれば、ステップ１６０２に進み、要求データがキャッシュメモリ（１４２）に存在するかどうか判定する。ステップ３９０１でＷＲＩＴＥ処理要求であれば、ステップ３９０３に進む。ステップ３９０３では、ホストプロセッサ（１０１）からの要求をそのままローカルディスクアレイ制御装置（３５０２，３５０３，３５０４，３４０５）に転送する。

【０２１４】図４２は、本実施例２の動作の一例を示している。

35

【０２１５】グローバルディスクアレイ制御装置（３５１３）から、データ４００２、入出力データ長４００３、入出力先頭アドレス４００４、仮想ドライブ識別子４００５、およびコマンド４００６が、一つのパケットとして、全てのローカルディスクアレイ制御装置（３５０２，３５０３，３５０４，３４０５）に転送される（４００１）。

【０２１６】ローカルディスクアレイ制御装置（３５０２，３５０３，３５０４，３４０５）での実行判定の結果、３５０２，３５０５では自ローカルディスクアレイ制御装置には関係のない入出力となりコマンドキャンセルを行なう。ローカルディスクアレイ制御装置（３５０３，３５０４）は、自ローカルディスクアレイ内の領域であると判断し、グローバルディスクアレイ制御装置（３５１３）から転送されたコマンドを解釈し実行する（４００７，４００８）。

【０２１７】こういった処理により、ローカルディスクアレイ制御装置（３５０２，３５０３，３５０４，３４０５）の接続性を良くすることが可能となる。つまり、グローバルディスクアレイ制御装置（３５１３）内では、入出力制御においてローカルディスクアレイ制御装置（３５０２，３５０３，３５０４，３４０５）を意識する必要がなくなった。従って、ローカルディスクアレイ制御装置を新たに追加しても、グローバルディスクアレイ制御装置（３５１３）の変更は必要ない。さらには、ローカルディスクアレイ制御装置も、制御メモリ（３５０７－３５１０）内のローカルディスクアレイ識別子を変更するだけで増設が可能となる。

【０２１８】

【発明の効果】本発明によれば、多数のディスク装置を接続しても、高速な入出力が可能であり、コストを上げることがなく、またディスク管理も容易で信頼性を低下させることもない。

【図面の簡単な説明】

【図１】本発明によるディスクアレイの一実施例の全体図を示す。

【図２】本発明による実施例１の動作概要を示す。

【図３】本発明による実施例１の動作概要を示す。

【図４】本発明による実施例１の動作概要を示す。

【図５】本発明による実施例１の動作概要を示す。

【図６】コマンド制御部の構成を示す。

【図７】制御メモリの構成を示す。

【図８】ローカルディスクアレイ管理テーブルの構成を示す。

【図９】仮想ドライブ管理テーブルの構成を示す。

【図１０】仮想ドライブの概要を示す。

【図１１】Ｉ／Ｏ管理リストの構成を示す。

【図１２】キャッシュ管理リストの構成を示す。

【図１３】コマンドキューリストの構成を示す。

【図１４】コマンド受付処理プログラムのフローを示す。

36

【図１５】終了処理プログラムのフローを示す。

【図１６】グローバルＲＡＩＤ制御部の構成を示す。

【図１７】Ｉ／Ｏ要求受付処理プログラムのフローを示す。

【図１８】データ配置制御プログラムのフローを示す。

【図１９】パススイッチの動作を示す。

【図２０】グローバルパリティ制御プログラムのフローを示す。

【図２１】グローバルパリティの生成概要を示す。

【図２２】コマンドパケットの流れを示す。

【図２３】タグ制御プログラムのフローを示す。

【図２４】データ回復制御プログラムのフローを示す。

【図２５】Ｉ／Ｏ要求終了プログラムのフローを示す。

【図２６】ローカルディスクアレイのＲＡＩＤ制御部の構成を示す。

【図２７】ローカルディスクアレイの制御メモリの構成を示す。

【図２８】ローカルディスクアレイのＩ／Ｏ要求終了プログラムのフローを示す。

【図２９】ローカルディスクアレイのデータ配置制御プログラムのフローを示す。

【図３０】ローカルディスクアレイのパリティ制御プログラムのフローを示す。

【図３１】ローカルディスクアレイのデータ回復制御プログラムのフローを示す。

【図３２】ローカルディスクアレイのコマンド受付処理プログラムのフローを示す。

【図３３】ローカルディスクアレイの終了処理プログラムのフローを示す。

【図３４】本発明によるデータ回復処理の動作を示す。

【図３５】本発明によるデータ格納例を示す。

【図３６】本発明によるデータ格納例を示す。

【図３７】本発明による実施例２の全体構成図を示す。

【図３８】実施例２のグローバルＲＡＩＤ制御部の構成を示す。

【図３９】実施例２の制御メモリの構成を示す。

【図４０】実施例２のコマンド受付処理プログラムのフローを示す。

【図４１】実施例２のデータ配置制御プログラムのフローを示す。

【図４２】実施例２の動作例を示す。

【符号の説明】

１０１…ホストプロセッサ、１０２…ディスクアレイ制御装置、１０８…インタフェース制御部、１３３，１３４，１３５，１３６…ディスク装置、１３７，１３８，１３９，１４０…ディスクドライブ、１０３…グローバルディスクアレイ制御装置、１０４，１０５，１０６，１０７…ローカルディスクアレイ制御装置、１４１…コマンド制御部、１４５…制御メモリ、１４２…キャッシ

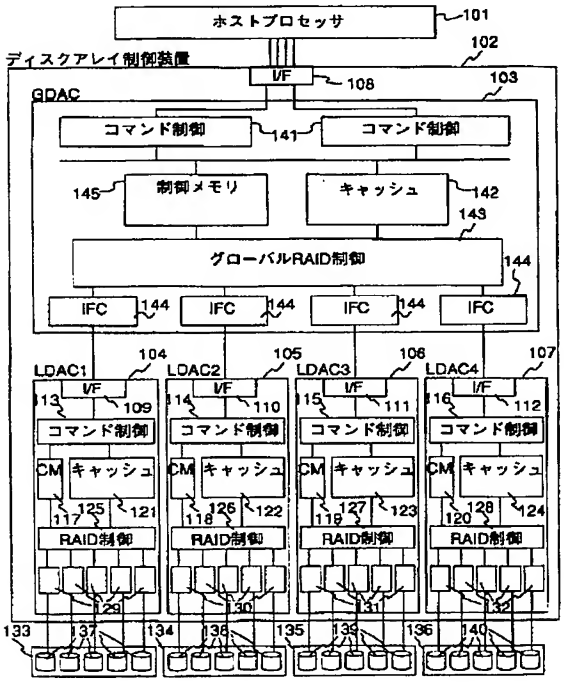
37

ュメモリ、１４３…グローバルＲＡＩＤ制御部、１４４…インタフェースコントローラ、４０２…プロセッサ、、５０２…仮想ドライブ管理テーブル、５０３…Ｉ／Ｏ管理リスト、５０４…ローカルディスクアレイ管理テーブル、５０５…キャッシュ管理リスト、５０６…コマンドキューリスト、４０５…コマンド受付処理プログ

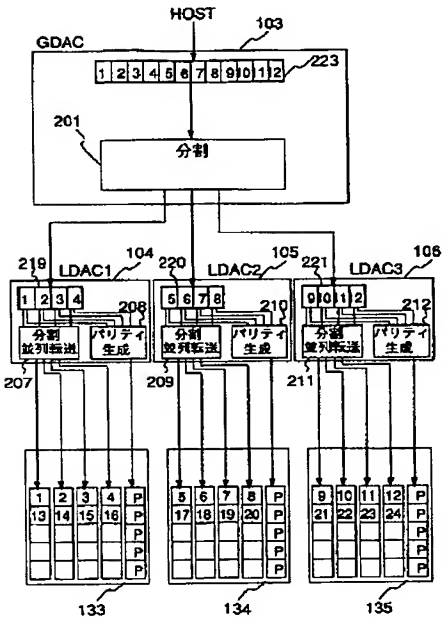
38

ラム、４０６…終了処理プログラム、１４０７…Ｉ／Ｏ要求受付処理プログラム、１４０８…データ回復処理プログラム、１４０９…Ｉ／Ｏ要求終了処理プログラム、１４１０…データ配置制御プログラム、１４１１…グローバルパリティ制御プログラム、１４１２…タグ制御プログラム、１４０５…パススイッチ。

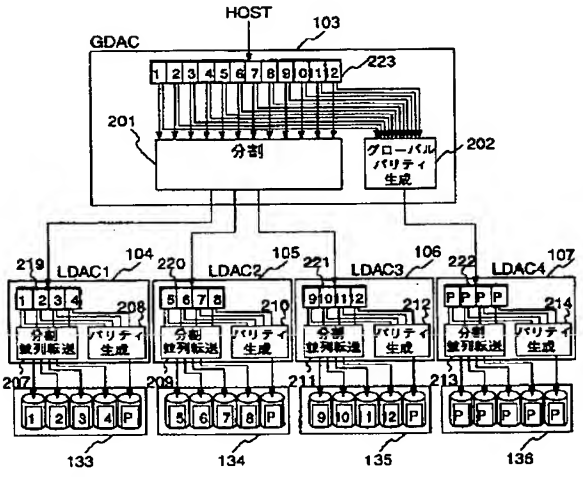
【図１】



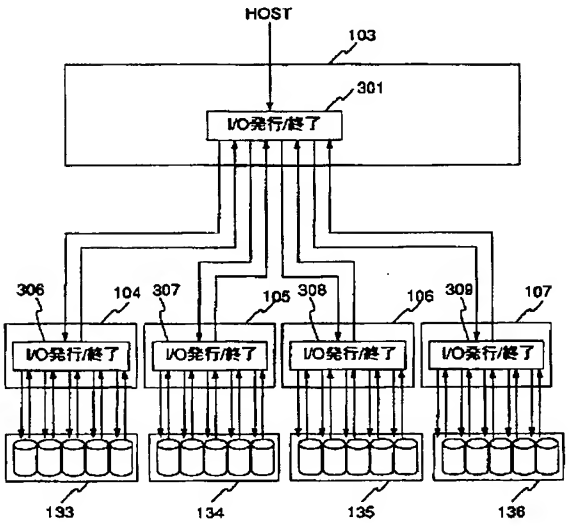
【図２】



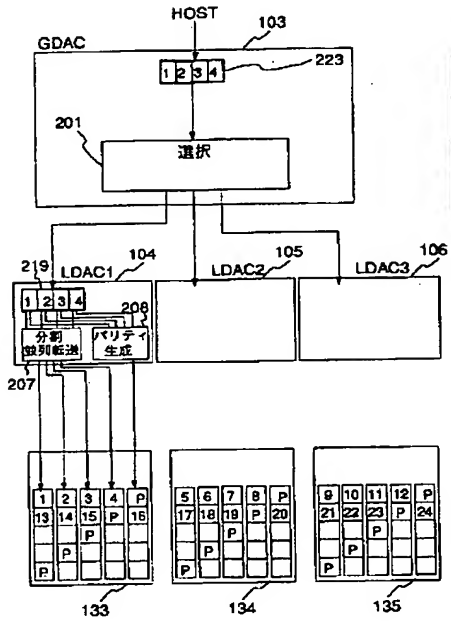
【図４】



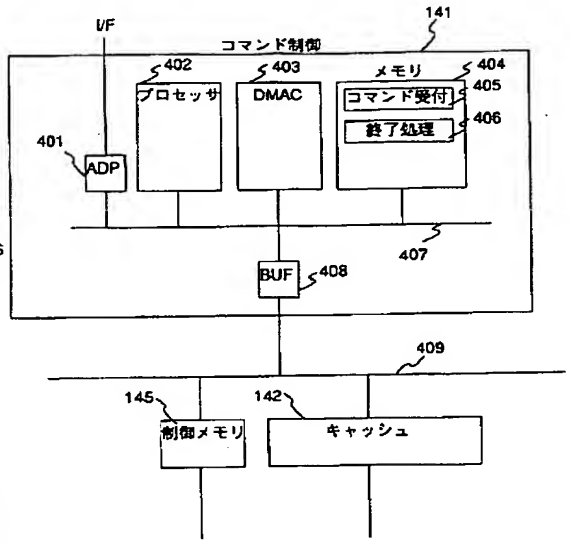
【図５】



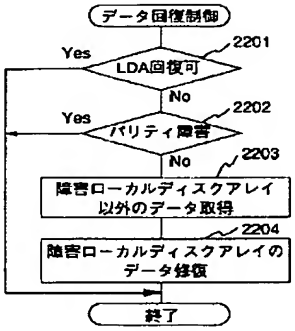
【図３】



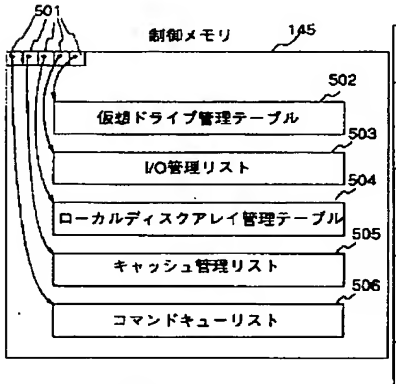
【図６】



【図２４】



【図７】



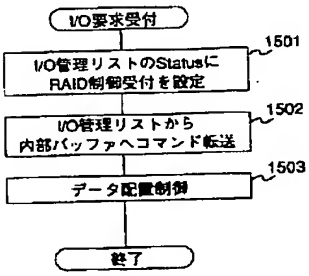
【図８】

ローカルディスクアレイ管理テーブル

LDAC	PART	PASS	DRV	RAID	PRTY	ST-S	CAP.	STATUS
LDAC1	1	1	5	5	1	1	1	Normal
	2	1	5	3	1	4	1	Normal
	3	1	5	0	1	1	2	Normal
LDAC2	1	2	5	5	1	1	1	Normal
	2	2	5	3	1	4	1	Normal
	3	2	5	0	1	1	2	Normal
LDAC3	1	3	5	5	1	1	1	Normal
	2	3	5	3	1	4	1	Normal
	3	3	5	0	1	1	2	Normal
LDAC4	1	4	5	5	1	1	1	Normal
	2	4	5	3	1	4	1	Normal
	3	4	5	0	1	1	2	Normal

601 602 603 604 605 606 607 608 609

【図１７】



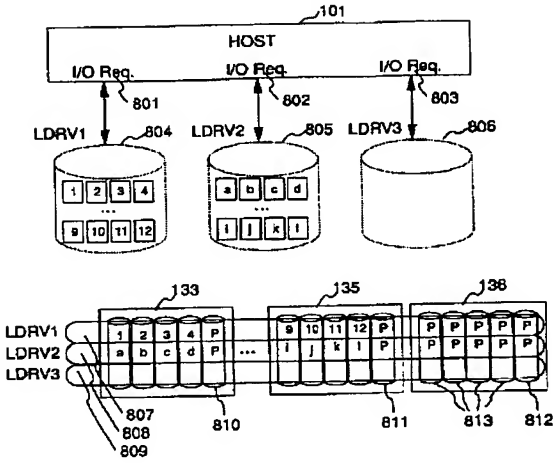
【図９】

仮想ドライブ管理テーブル

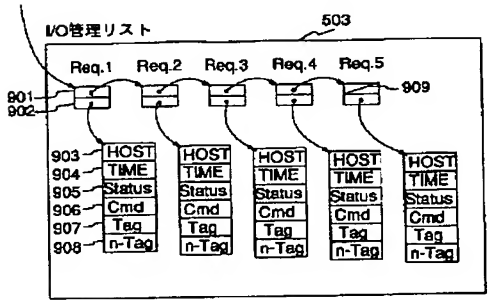
ドライブID	ローカルディスクアレイ領域 PART	PART	PART	PART	Cap.	グローバルRAID	PRTY	ST-S	STATUS
VOL1	1	1	1	1	4	0	1	3	Normal
VOL2	2	2	2	2	4	3	1	18	Normal
VOL3	3	3	3	3	8	5	1	4	Normal

701 702 703 704 705 706 707 708 709 710

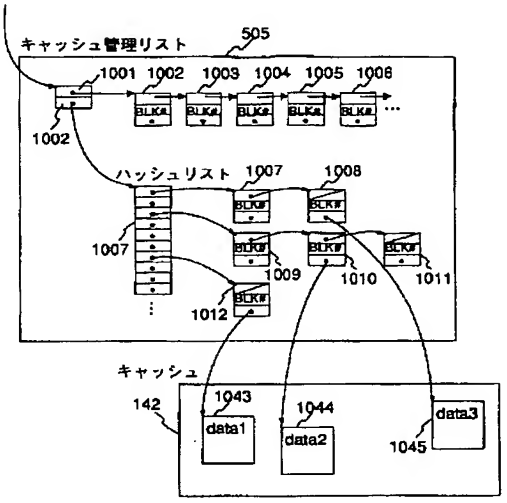
【図１０】



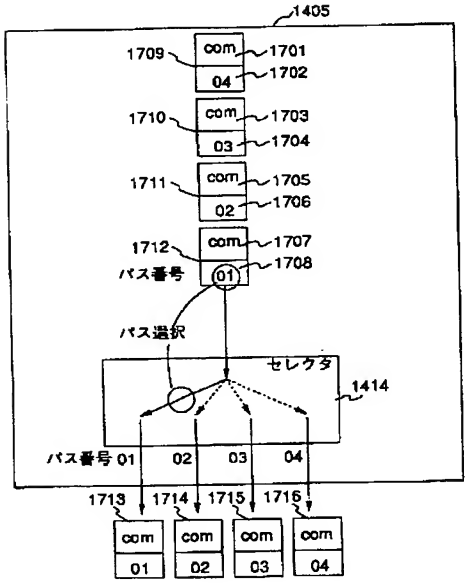
【図１１】



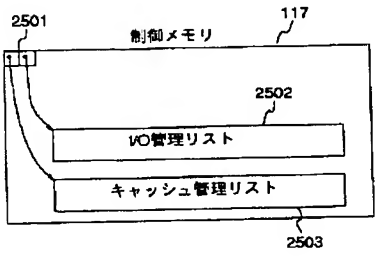
【図１２】



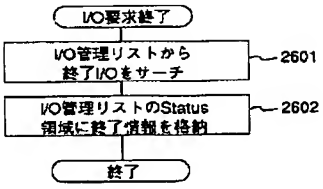
【図１９】



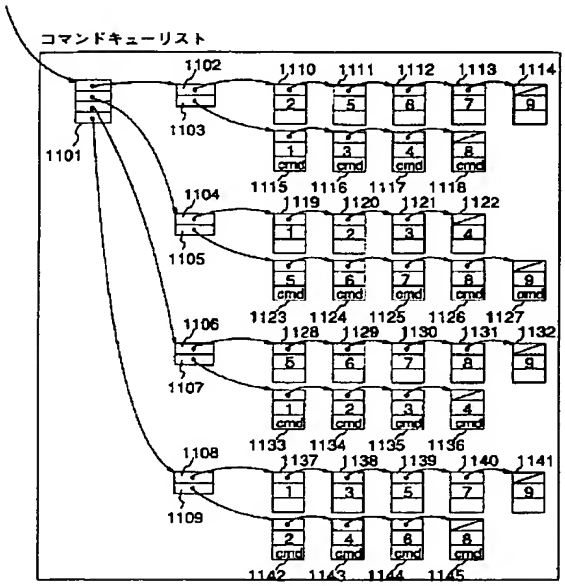
【図２７】



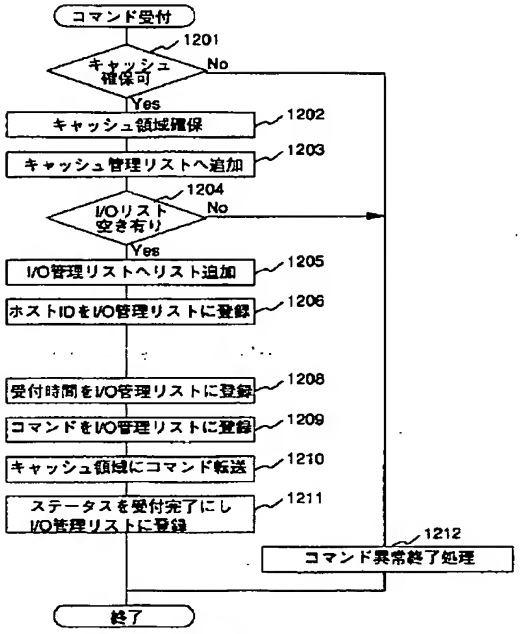
【図２８】



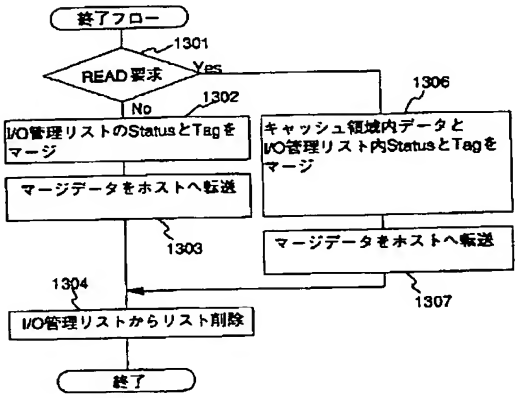
【図１３】



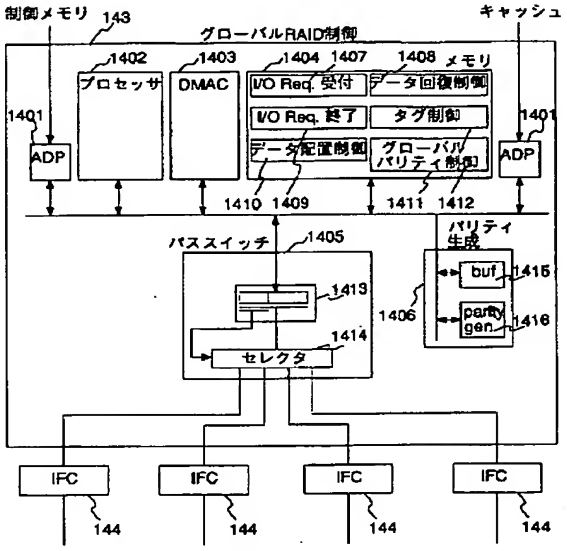
【図１４】



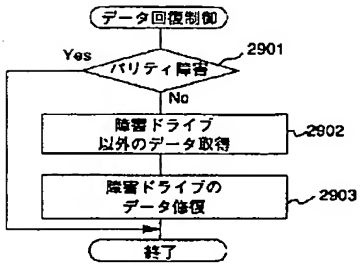
【図１５】



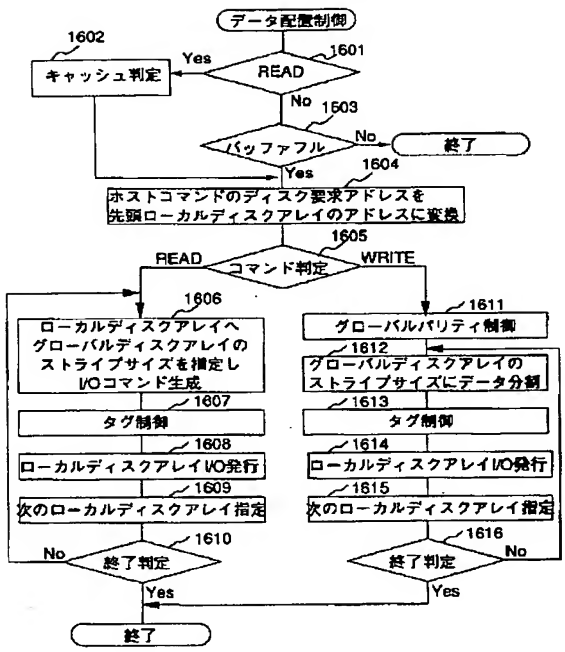
【図１６】



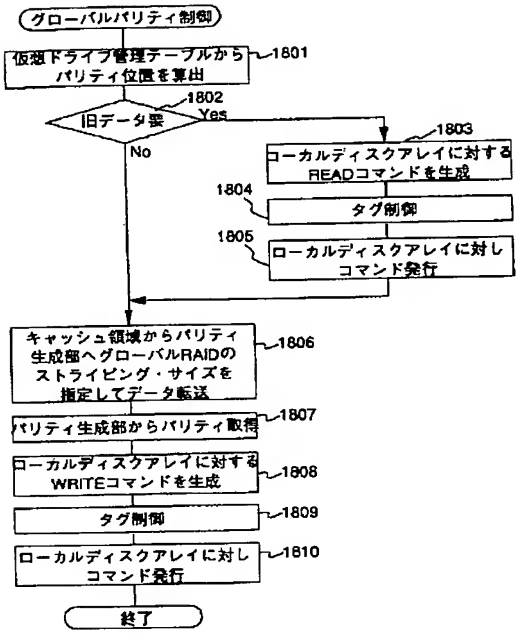
【図３１】



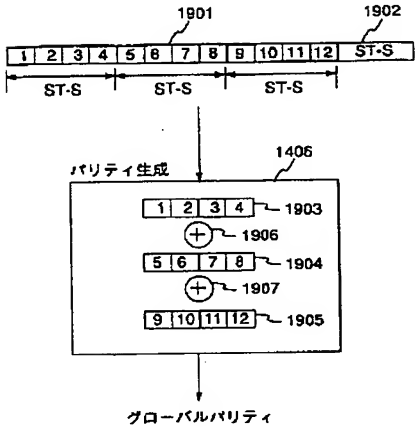
【図１８】



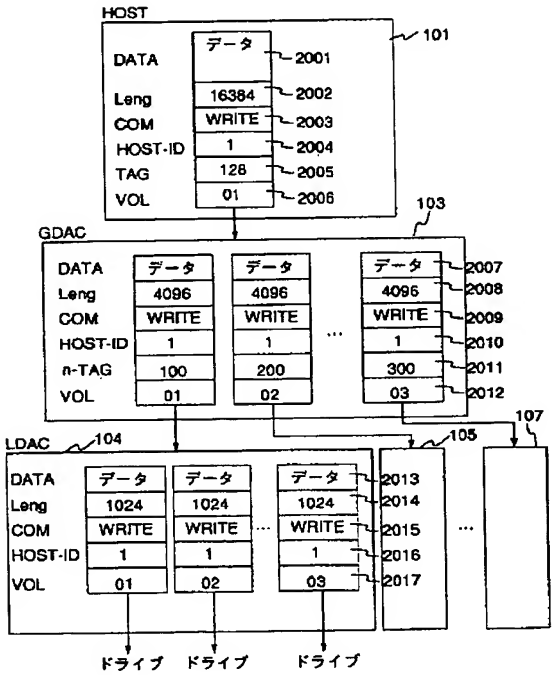
【図２０】



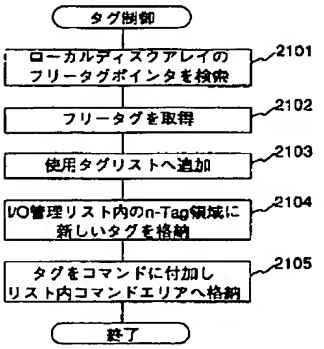
【図２１】



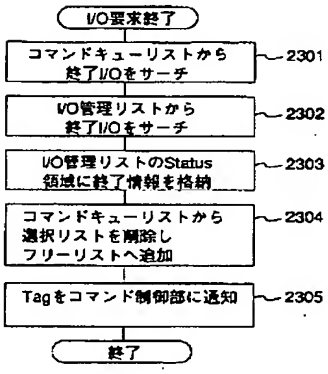
【図２２】



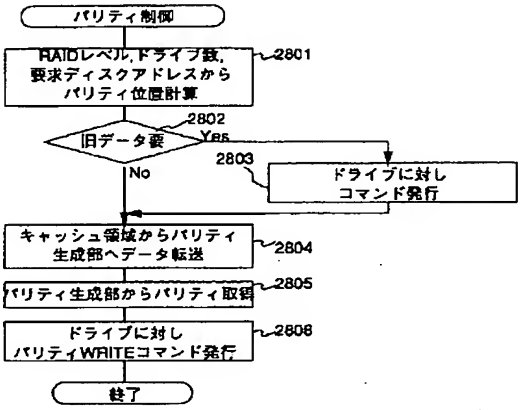
【図２３】



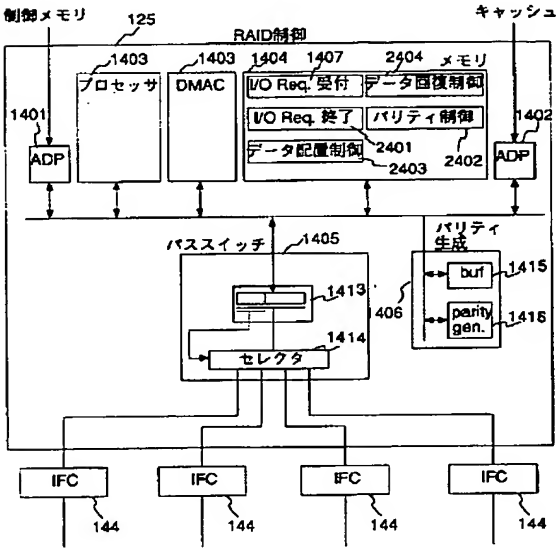
【図２５】



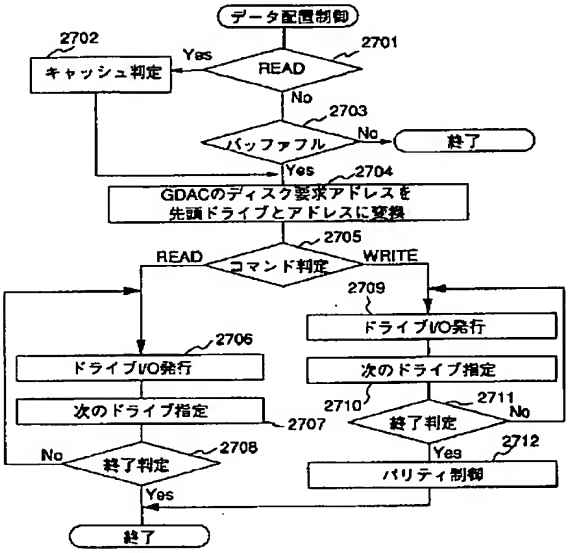
【図３０】



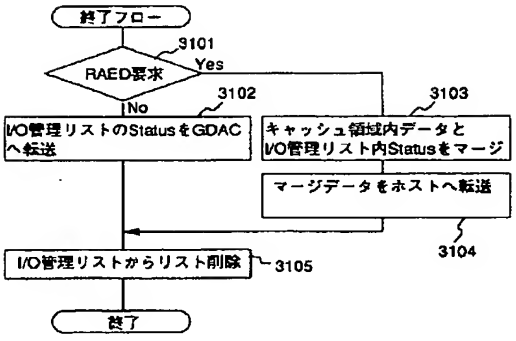
【図２６】



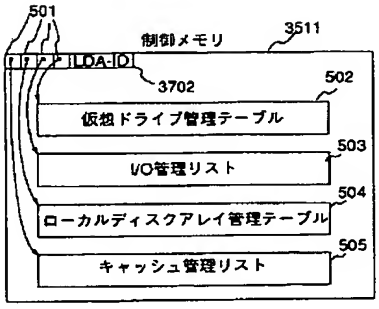
【図２９】



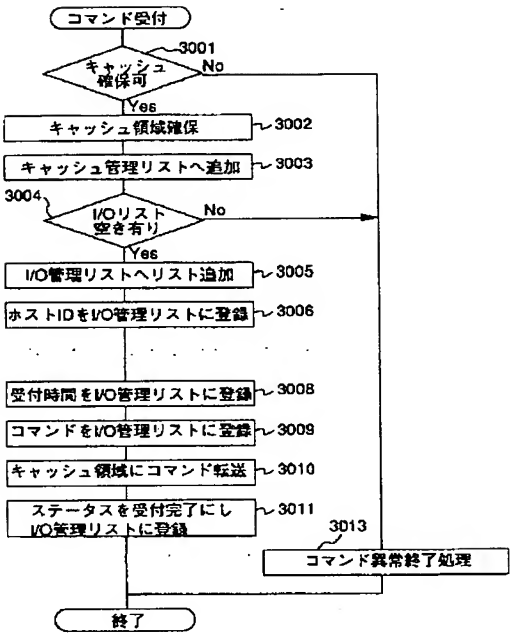
【図３３】



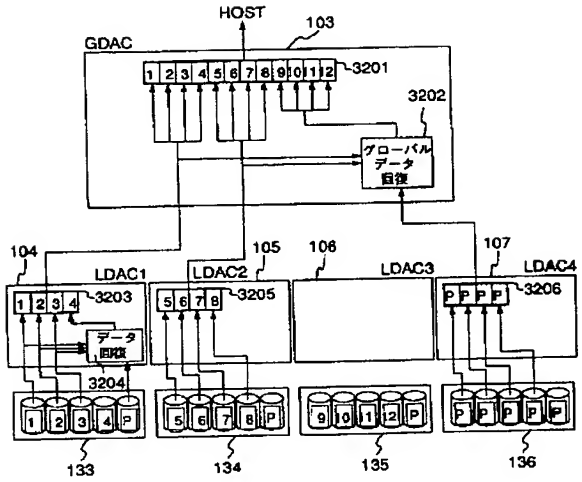
【図３９】



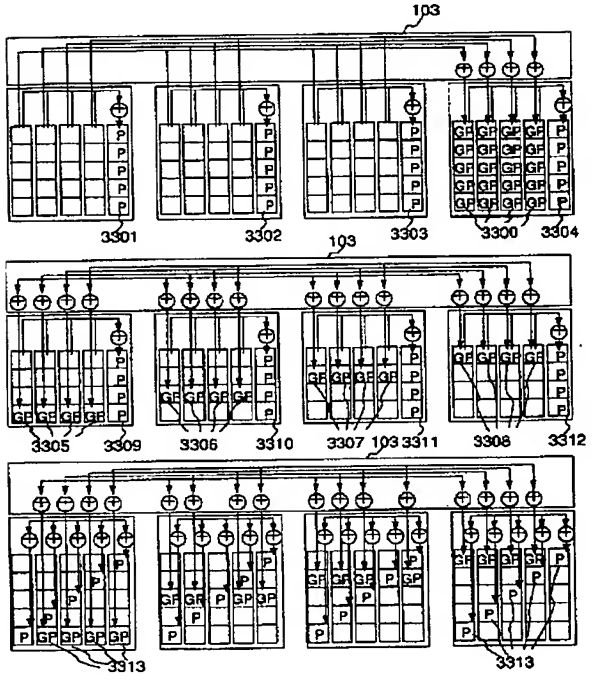
【図３２】



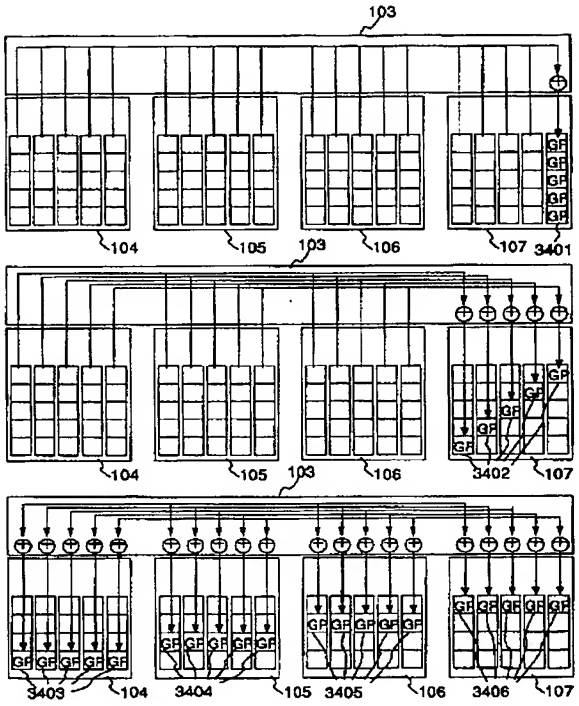
【図３４】



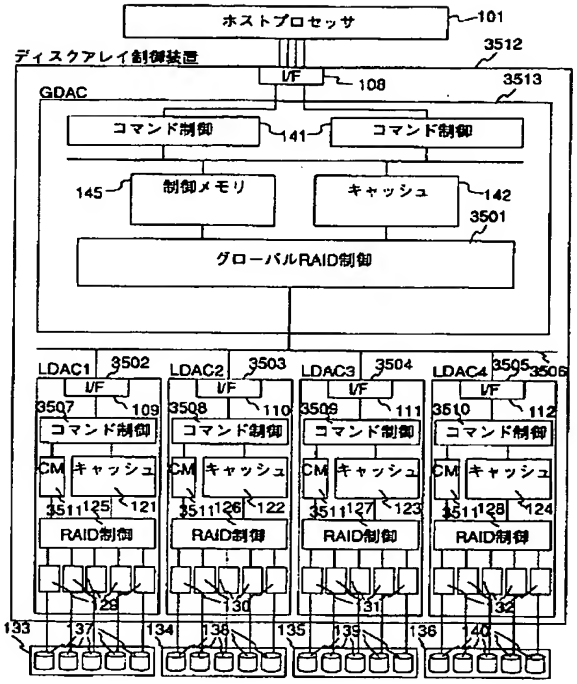
【図３５】



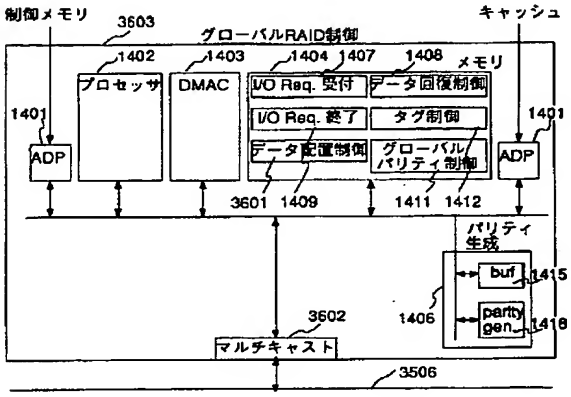
【図３６】



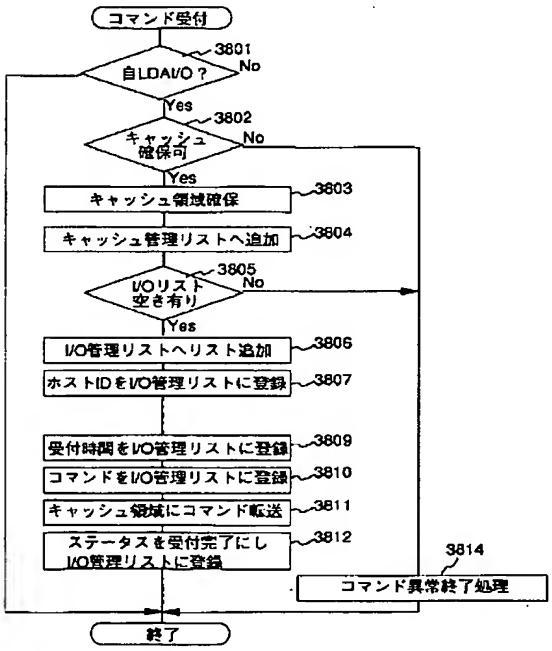
【図３７】



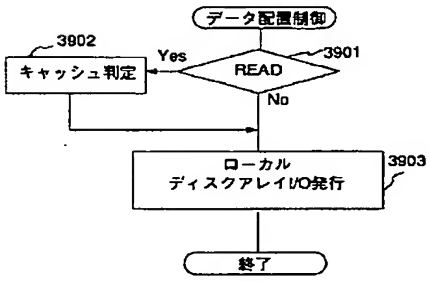
【図３８】



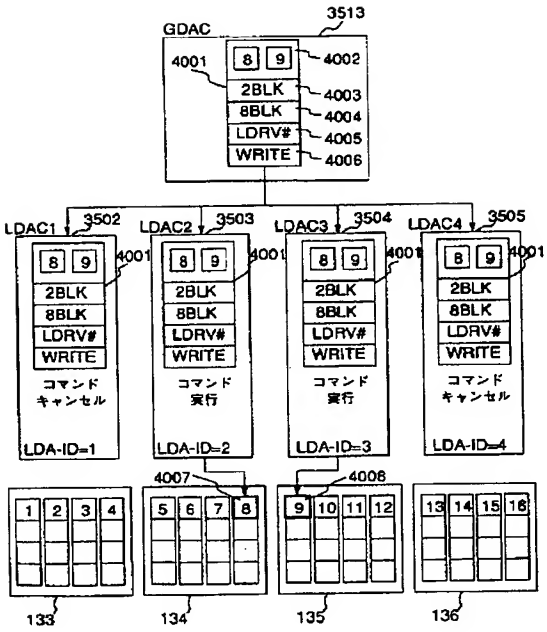
【図４０】



【図４１】



【図４２】


GDAC
3513
4001
8
9
4002
2BLK
4003
8BLK
4004
LDRV#
4005
WRITE
4006
LDAC1
3502
LDAC2
3503
LDAC3
3504
LDAC4
3505
コマンド
キャンセル
コマンド
実行
コマンド
実行
コマンド
キャンセル
LDA-ID=1
LDA-ID=2
LDA-ID=3
LDA-ID=4
4007
4008
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
133
134
135
136